Project Title: Assessing the effectiveness of environmentally-sensitive design (ESD) for achieving stormwater management objectives in the Upper Little Patuxent River Watershed, Howard County, MD

Principal Investigator: Keith N. Eshleman, Ph.D.
University of Maryland Center for Environmental Science
Appalachian Laboratory
Frostburg, MD 21532

Final Report to: Chesapeake Bay Trust
Restoration Research Program

Date: December 31, 2023

Executive Summary

This report (and the accompanying database) represents the culmination of a Chesapeake Bay Trust Restoration Research Program project to assess the effectiveness of environmentally-sensitive design (ESD) for achieving stormwater management objectives in the Upper Little Patuxent River Watershed (ULPR) watershed in Howard County, Maryland. The overall objective of the project was to assess the spatially-aggregated effectiveness of ESD best management practices (BMP's, e.g., rain gardens, dry wells, bioretention facilities, etc.) in addressing four primary stormwater management goals at the watershed scale: (1) protection of receiving surface waters from nonpoint source pollution; (2) attenuation of stormwater discharge peaks; (3) diminishment of stormwater runoff volumes; and (4) enhancement of the recharge/discharge behavior of shallow groundwater that sustains stream baseflows. While we attempted to employ a true BACI design using paired watersheds to address the project objective, the study was hampered by the fact that the "pre-ESD" period of six months was necessarily shorter than a full water year—so the study is perhaps better described as a "quasi-BACI" design. Notwithstanding this shortcoming, we were able to successfully monitor two small subwatersheds of ULPRB over a 3.5-year period (water years 2020 through 2023): 1) Plumtree Branch (PLBR) encompassing the headwaters of a ULPRB tributary draining a small (2.15 km^2), highly suburbanized subwatershed that includes much of the older Valley Mede neighborhood located in the north-central part of Ellicott City; and 2) an unnamed tributary to ULPRB draining a small (0.80 km^2), developing portion of the basin in the western part of Ellicott City, MD (UTLP). Most of the PLBR watershed was developed prior to 1984 when stormwater regulations went into effect in Maryland, while UTLP was mostly developed post-1984; the most recent development in UTLP encompassing about 25 acres (0.10 km^2) of previously-forested land relied exclusively on implementation of ESD BMP's (i.e., 58 dry wells, nine bioretentions, and 11 micro-bioretentions to control stormwater impacts. We successfully characterized stormflow responses to *all* 101 discrete rainfall events over the course of the project (snowmelt events were not characterized). We were also able to separate stormflow hydrographs from most of the 101 events (including 60 common events) using 5-min specific conductance (SC) data and calibrate a recursive digital filter (RDF) model of hydrograph separation. Finally, we analyzed nearly 1300 discrete water samples collected during both baseflow and stormflow conditions for 18 different constituents and used an empirical load model (LOADEST) to estimate annual loads.

Despite the very short (~6-months) pre-treatment period, the monitoring data provided very strong support for our hypothesis that ESD implementation in the UTLP watershed progressively reduced total runoff. Total runoff was shown to decrease by about 42% relative to PLBR, but this is likely an overestimate of the overall effect given the estimated 17% difference based on data from the pre-ESD period. The difference for the pre-ESD period could be explained by one or more factors including uncertainties in delineating the watersheds and a real difference in evapo-transpirative demand due to greater forest cover in UTLP. Therefore, our best estimate for the ESD effect on total runoff was a decrease of about 25%. This reduction in total runoff is actually about a factor of two larger than the percentage of the UTLP watershed that was disturbed (~13%) during the most recent development sequence—suggesting that ESD has had an outsized impact on total runoff. There are several possible mechanisms that could explain the disproportionate effect of ESD BMP's: 1) enhanced groundwater recharge associated with both the bioretentions and the dry wells; 2) increased evapotranspirative losses of water either in the bioretention facilities themselves or, more likely, from areas to which the bioretentions discharged water; and 3) off-watershed diversion of stormwater (e.g., to a regional stormwater pond).

Statistically-significant changes in storm event runoff, runoff ratios, peak hourly event runoff, and maximum new water contributions (quantified as median differences between PLBR and UTLP) between the two study periods are at least qualitatively consistent with the observed decrease in total runoff and demonstrate *en masse* the robustness of the hydrologic effects of ESD implementation. Each of these effects is consistent with another effect that we managed to detect using the optimized RDF approach underpinned by an extensive set of chemical hydrograph separations: a monotonic increasing trend in baseflow index (BFI) in UTLP with no trend noted for the control watershed (PLBR). It is important to note here that we did *not* observe a dramatic increase in baseflow volumes in the UTLP watershed. In fact, annual baseflow runoff did not change much during the study: baseflow volumes were particularly stable in full water years 2021 through 2023. Conversely, direct runoff volumes declined by about 50% from 2021 (110mm) to 2021 (55mm)—dominating the observed BFI signal. At least a portion of the observed decline in direct runoff (and increase in BFI) could have been due to lower precipitation in water years 2022 and 2023 relative to 2021 (as we observed at PLBR), but we consider it unlikely that hydroclimatic variability alone could explain this phenomenon.

While the hydrologic metrics provided strong support for changes attributable to ESD implementation at the watershed scale, the water quality results were much less conclusive. We were unable to detect any significant changes in water quality during the 3.5-year project. Baseflow water quality proved particularly stable in both watersheds—presumably due to long subsurface residence times that effectively slows the pace of geochemical evolution and pollutant mobility. Stormflow was obviously quite dynamic and clear patterns were observed using event mean concentrations as integrative indices of stormwater quality, but extreme hydroclimatic variability apparently overwhelmed the impacts that ESD implementation might have had on water quality and loads. One problem that we identified earlier was the extremely short pre-ESD period. This problem turned out to be even more consequential for the water quality part of the study than for the hydrologic component. We simply were unable to characterize water quality variations for a sufficient number of storms during the pre-ESD period of the project to provide a tighter baseline characterization for the latter part of the study. A second issue is that despite strong support in the literature, the loading model (LOADEST) that we chose to work with may not be (in hindsight) the best choice for watersheds that are exhibiting non-stationary hydrologic behavior. For future load modeling work, we propose using the newer WRTDS load model that may be inherently better suited for addressing water quality variations in watersheds disturbed by urbanization or other rapid land use changes.

While we obviously could not go back and increase the length of the monitored pre-treatment period, we *are* committed to continuing the monitoring effort in both watersheds for several additional years (i.e., collecting equivalent data through a "post-ESD" period). Assuming no other changes in land use in the study watersheds occur in the immediate future, our expectation is that data for the post-ESD period would strongly reinforce the hydrologic and water quality changes that emerged during the current project. Future monitoring would build on the data foundation provided by the current project, as well as allow us to test some other data analysis methods. Fortunately, we were able to secure additional funding from Chesapeake Bay Trust Restoration Research Program that will allow for the overall project to continue through water year 2025. We will be relying on the monitoring systems that remain in place in these Howard County watersheds to extend the project into the post-development phase.

Introduction

Urbanization of watersheds has contributed to a variety of hydrological and biological changes in streams that are often included into what aquatic ecologists have termed the "urban stream syndrome" (Walsh et al. 2005). Perhaps the most significant hydrological changes involve the alteration of flow regimes (e.g., increases in storm event runoff volumes and peak runoff) and flowpaths (e.g., greater contributions of overland runoff) caused primarily by increasing watershed imperviousness such as roof surfaces, roads, sidewalks, and parking areas (O'Driscoll et al. 2010; Hopkins et al. 2015). The widespread use of conventional ("gray") stormwater management practices (i.e., concrete curbing, grated inlets, and underground culverts) that facilitate the rapid conveyance of urban runoff into streams has undoubtedly exacerbated the urban stream syndrome. Newer "green" stormwater infrastructure (i.e., rain gardens, dry wells, and bioretention facilities that are considered part of environmentally-sensitive design, ESD) that is designed to temporarily detain and/or retain stormwater runoff has the potential to mitigate some of the negative impacts of urbanization, however (Scarlett et a. 2018). The effectiveness of these ESD practices can be quantified by the extent to which the following stormwater management goals are achieved (Davis et al. 2009): (1) protection of receiving surface waters from nonpoint source pollution; (2) attenuation of stormwater discharge peaks; (3) diminishment of stormwater runoff volumes; and (4) enhancement of the recharge/discharge behavior of shallow groundwater that sustains stream baseflows.

While a reasonably extensive literature has rapidly evolved to quantify the effectiveness of GSI at the site or project scale (Dietz 2007, Barrett 2008, Liu et al. 2014, Kratky et al. 2017), there have been relatively few carefully-designed experimental studies that have demonstrated the effectiveness of ESD practices at the watershed scale (Rose and Peters 2001, Hood et al. 2007, Selbig and Bannerman 2008, Bedan and Clausen 2009, Schuster and Rhea 2013, Loperfido et al. 2014, Jarden et al. 2016; Scarlett et al. 2018). The cited studies have addressed this problem at only seven different urban locations in the US (Atlanta GA, Waterford CT, Cross Plains WI, Cincinnati OH, Washington DC metro area, Parma OH, and Charlotte NC) and none of these studies addressed all four stormwater management objectives. Moreover, only one of the seven studies (Bedan and Clausen 2009) was able to demonstrate the effectiveness of using ESD BMPs in achieving water quality goals (relative to conventional stormwater infrastructure), so another comprehensive study of ESD effectiveness that

considers both water quality and quantity at the watershed scale would represent a valuable addition to the scientific and management literature.

Three of the eight cited studies (Bedan and Clausen 2009, Schuster and Rhea 2013, Jarden et al. 2016) employed the well-accepted "before-after, control-impact" (BACI) design in which the actual observed impacts following development/treatment are quantified relative to predicted impacts based on a pre-development calibration with a control system—thus allowing for explicit tests of causality as the method effectively corrects for year-to-year differences in precipitation and other hydroclimatic variables. It is important to emphasize that the BACI design is really only tractable for cases where development occurs very rapidly and/or long-term funding has been secured to support monitoring throughout pre-development, during-development, and post-development conditions (i.e., often 5 to 10 years or even longer). The alternative to a BACI design is one in which the researcher attempts to carefully "control" for non-impact related response differences through careful watershed selection—as was the case for most of the cited studies.

Project Objective

This report (and the accompanying database) represents the culmination of a Restoration Research Program project entitled "Assessing the effectiveness of environmentally-sensitive design (ESD) for achieving stormwater management objectives in the Upper Little Patuxent River Watershed, Howard County, MD" that was funded by Chesapeake Bay Trust in 2019. The overall objective of the project was to assess the spatially-aggregated effectiveness of ESD BMPs in addressing the four primary stormwater management goals at the watershed scale through a comparative (i.e., "paired") experimental study implemented in the Maryland Piedmont.

Experimental Design

Initially, we proposed a strictly comparative field study of two Ellicott City (MD) watersheds both with relatively stable land use, but differing primarily with respect to the dominant stormwater management system in place: 1) one watershed with conventional management; and 2) one watershed with ESD. Unfortunately, after instrumentation in the watershed with conventional management had already been installed in late 2019, we were unable to get permission to access and establish a stormwater monitoring station in the second watershed—forcing us to consider other study design options. After evaluating several options, including consideration of logistical and

watershed access issues, we proposed a hybrid (i.e., quasi-BACI) design that would include monitoring of the mostly developed watershed with conventional stormwater management (the "control" watershed) and an adjacent watershed that was in the process of being developed using ESD stormwater practices (the "impact" watershed). It was determined that a quasi-BACI design could also be valuable in determining whether a watershed being developed with ESD could be shown to be progressively achieving stormwater management goals (relative to a "control" watershed that was developed using conventional stormwater infrastructure).

With this alternate experimental design, we could thus focus the project on evaluating differences (between the two watersheds) in stormwater metrics over time, including variations in the metrics between three time periods: 1) a "pre-ESD" period; 2) a "during ESD" implementation period; and 3) a "post-ESD" period. While not the primary determinant, this hybrid design also made sense logistically. Howard County had ownership of land parcels in both watersheds that were deemed suitable for our stormwater monitoring stations and the county was willing to grant us legal right-of-entry for the life of the research project. Permits to install and operate field instrumentation at the outlets of the two watersheds (both located on county property) were obtained in the form of memoranda of understanding (MOU) between the University of Maryland Center for Environmental Science (UMCES) and Howard County.

Unfortunately, literally within 24 hours after monitoring equipment had been successfully installed in the "impact" watershed and "paired" monitoring was underway in both watersheds, the onset of the COVID-19 pandemic would rear its ugly head and cause yet another project design change. The pandemic dramatically slowed the pace of development and residential construction, so that we were eventually forced to abandon the possibility of collecting data during the "post-ESD" period (even after we voluntarily extended the field monitoring for an additional year at no additional cost to the sponsors). The present report thus includes only data collected and analyzed for the "pre" and "during" periods of ESD implementation, although "post-ESD" monitoring commenced in October 2023 as part of a follow-on project funded by CBT.

Study Watersheds and Land Use Change

Virtually all of the data necessary for achieving the project objectives were obtained through hydrologic monitoring of two small subwatersheds of the Upper Little Patuxent River Basin (ULPRB) in the Piedmont region of Howard County, Maryland: 1) Plumtree Branch (hereafter denoted PLBR) encompassing the headwaters of a ULPRB tributary draining a small (2.15 km^2), highly suburbanized subwatershed that includes much of the older Valley Mede neighborhood located in the north-central part of Ellicott City; and 2) an unnamed tributary to ULPRB draining a small (0.80 km^2), developing portion of the basin in the western part of Ellicott City, MD (hereafter denoted UTLP for Unnamed Tributary Little Patuxent; Figure 1). According to USGS StreamStats, the PLBR watershed contained 28.1% impervious surface area (based on 2011 NLCD data), while the UTLP watershed contained 17.8% imperviousness (Table 1); the watersheds are spatially contiguous and are both bisected by six lanes of Interstate 70. Much of the PLBR watershed was developed in the 1960's through 1980's and can be characterized as single family development; residential lots are relatively large (~0.5 acres) by contemporary standards. Interstate-70 bisects the northernmost portion of the watershed. With the exception of a large parcel of land located in the most northwest quadrant of the watershed, the watershed is nearly fully built-out. During the study period, we did not observe any major development occurring in the PLBR basin, so for our purposes it represents a developed control watershed with stable land use. According to a 2019 land use

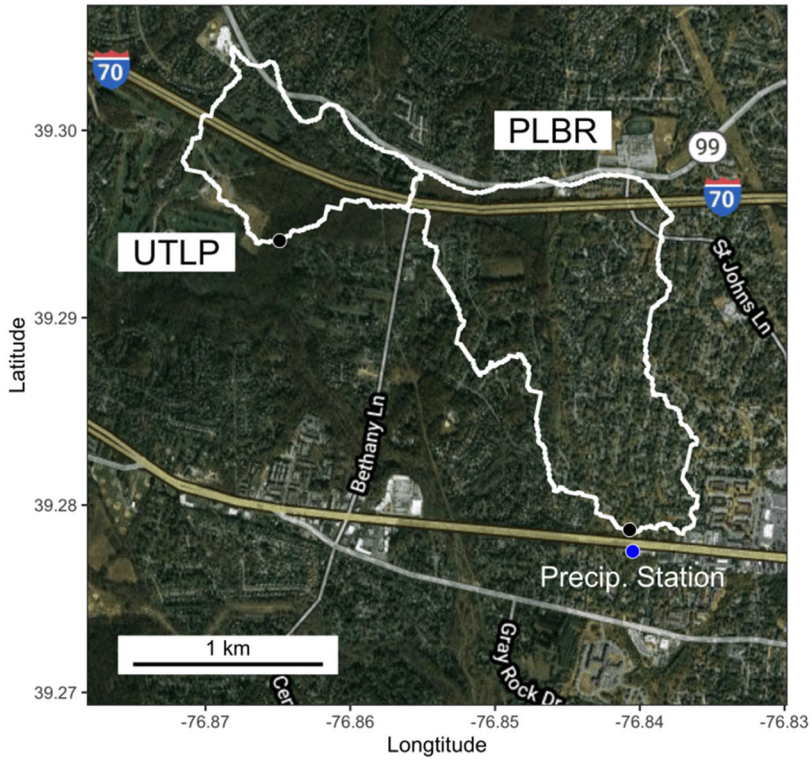

Figure 1. Map of study watersheds in Howard County, Maryland with stormwater monitoring stations (black circles) and precipitation station ("Kappa"; blue circle) shown. An auxiliary precipitation station ("Delta") located several kilometers west of the watersheds on US Rte. 40 near the intersection with Marriottsville Road is not shown.

analysis conducted by Biohabitats, Inc., more than 80% of the PLBR watershed was developed prior to 1984 when stormwater controls were first mandated, although there are five stormwater ponds that provide some degree of stormwater detention (Hribar and Lyons 2017). Quite a few homes in the PLBR watershed, particularly those built within the floodplain, have been affected by repeated flooding in recent decades—forcing the county to address stormwater issues in the basin (Biohabits, Inc. 2019). This includes a small, county-owned parklet on Longview Drive that was established after several homes were condemned and torn down in the aftermath of the most recent urban flooding that occurred in 2016 and 2018; this parklet serves as the location of the PLBR stormwater monitoring station established in 2019.

Table 1. Characteristics of the study watersheds based on land use data (NLCD 2011) obtained via StreamStats (https://streamstats.usgs.gov/ss/).

Subwatershed	Area (km^2)	Impervious (%)	Forest (%)	Developed (urban) %
UTLP	0.80	17.8	37.5	60.6
PLBR	2.15	28.1	17.7	76.8

Development in the adjacent UTLP watershed began in the 1980's and, prior to 2016, virtually all of the residential development was located north of bisecting I-70; building lots in this portion of the UTLP watershed are smaller than in PLBR (~0.25 – 0.33 acres), however. The development is also served by three stormwater ponds (two wet ponds and a dry pond; Table A1). Analysis of Google Earth imagery suggests that newer development south of I-70 likely began in late 2015 or early 2016

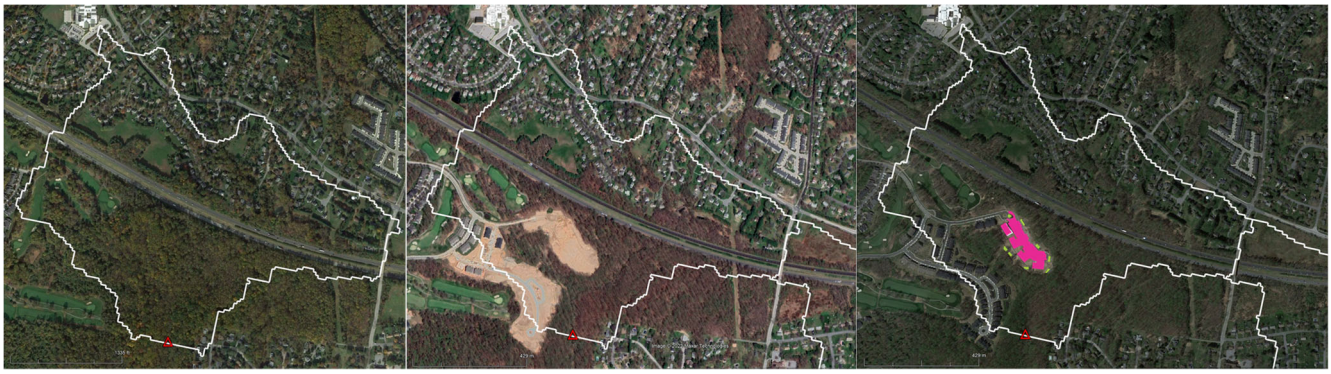

Figure 2. Recent development and GSI implementation trajectory in the UTLP watershed based on satellite imagery obtained from Google Earth for three dates: May 2015 (left); April 2020 (center); and April 2022 (right). The April 2022 panel includes an overlaid illustration of the active permitted development and GSI implementation in the watershed (pink polygons are impervious areas and green polygons are bioretention facilities that were under construction). The red triangle in each panel indicates the location of the UTLP stormwater monitoring station.

including an extension of Resort Road, construction of Vardon Lane, and associated forest clearing and mass grading for road construction, utilities, and residences (Figure 2). By spring 2020 when our stormwater monitoring equipment was installed in the UTLP watershed, the imagery shows that: 1) the remaining forest clearing for the latest phases of development, as well as much of the mass grading, had been completed; 2) four multi-family townhomes were either under construction or recently completed; 3) several other streets (i.e., Puccini Lane, Rossini Lane, and Vivaldi Lane) were under construction (several to the point of the first layer of asphalt). Sediment traps had also been constructed by this point and the entire disturbed area had been surrounded by double silt fencing to control erosion and sedimentation (E&S). Using the "ruler" tool in Google Earth and overlaying the watershed boundary obtained from StreamStats, we estimated that the total area of disturbance for the newest development in the UTLP watershed was about 25 acres (0.102 km^2 or 12.7% of the watershed area) based on the April 2020 imagery (Figure 2). Combining this disturbed area with the impervious area based on the 2011 imagery (0.142 km^2), the estimated combined area of disturbance plus impervious surface was 0.244 km^2 at the onset of the monitoring program.

Anecdotally, we observed relatively little progress in development over the next six months due to the pandemic, so we feel justified in considering the period from April through September 2020 (i.e., the last half of the 2020 water year) as the "pre-ESD" period. For interpretive purposes, however, it's important to recognize that the "pre-ESD" condition is definitely not a "pre-construction" period. Rather, the fact that a significant portion of the UTLP watershed had been cleared and was

Figure 3. Photos of ESD stormwater BMPs that were installed in the UTLP watershed during the monitoring project.

undergoing development suggested to us that we would likely observe some of the worst hydrological impacts of residential development during this period. Our hypothesis was that as the new development was completed and this portion of the watershed transitioned to a stable development served by ESD BMPs, we would observe a gradual decline in the most deleterious impacts.

While monitoring for this phase of the overall project was completed by 1 October 2023 (i.e., by the beginning of the 2024 water year) as construction effectively ended, we do not yet have an image showing the completed development; for this reason, we combined our analysis of the most recent image from April 2022 and development plans obtained from Howard County to estimate UTLP watershed imperviousness at the end of the "during-ESD" implementation phase. Using this method, we estimated a watershed imperviousness of 23.1%--very close to the 28.1% value for PLBR (Figure 2). In the three-year period from October 1 2020 through October 1 2023, ESD was gradually implemented as part of the new construction so we refer to this period as the "during-ESD" period. During this period, sediment traps that were being used as E&S controls were effectively converted to bioretentions or mico-bioretentions (Figure 3) and dry wells were constructed to control roof runoff impacts from many of the residential units. Based on the development plans, stormwater generated by the newly constructed impervious surfaces is treated by nine bioretentions (F-6), 11 micro-bioretentions (M-6), and 58 dry wells (M-5) distributed across the basin (Table A1).

Site Instrumentation

A stilling well (equipped with a steel streamwater intake pipe) and an instrument shelter housing a Unidata Model 6541 digital water level recorder (DWLR) were installed on the west streambank at the outlet of PLBR (Figure 1) on November 19, 2019 to provide continuous, 5-min water level data; a non-recording staff gage in the gage pool adjacent to the shelter was installed at the

Figure 4. Photo showing stormwater monitoring station at PLBR; automatic sampler and Aquatroll 500 not visible.

Figure 5. Photo showing stormwater monitoring station at UTLP; automatic sampler and Aquatroll 500 not visible.

same time to provide a visual reference of gage height against which the DWLR could be calibrated; in practice, no adjustments were needed. A separate PVC conduit between the stream and the shelter was also installed to accommodate an InSitu Aquatroll 500, a remote "tube" for powering the Aquatroll 500 and telemetering 5-min water level, water temperature, specific conductance (SC), and turbidity data to a project database on the HydroVu server (Figure 4); a portable, programmable Sigma sequential water sampler (not shown) was used to collect discrete streamwater samples for subsequent laboratory analysis of an extensive suite of water quality constituents (see Laboratory Methods section).

A comparable stormwater monitoring station was installed on the east bank of UTLP (Figure 1) on March 8, 2020; due to the fact that the streambed at the UTLP station is on a bedrock outcrop, we were unable to install a staff gage, so we used the top of the steel intake pipe as a reference for making manual gage height readings (Figure 5); as at PLBR, no adjustments of the DWLR were needed. Two unheated tipping bucket rain gages were used to characterize (15-min) rainfall at the study watersheds (Figure 1), although it quickly became apparent that the relatively high spatial variability in rainfall between the watersheds necessitated that we explore the use of WSR-88D (i.e., NEXRAD) Level 3 rainfall data as a way of estimating areal storm rainfall (see Field Methods section). The field instrumentation, combined with ancillary other field and laboratory measurements, allowed us to generate a plethora of both high frequency and moderate frequency data for characterizing the stormflow responsiveness of the two watersheds that is discussed in subsequent sections of the report.

Research Hypotheses

Given how the recent development occurred in UTLP since data collection began in March 2020, we made the *post hoc* decision to consider data collected in the 2020 water year (i.e., the period from early March through September 2020) representative of a "pre-ESD" period, while data collected

subsequently (i.e., from October 2020 through September 2023) was considered representative of a "during-ESD" implementation period. This classification—albeit imperfect—was necessary to support the hypothesis testing that would be used to evaluate hydrologic and water quality changes in the UTLP watershed and their relationships, if any, to ESD implementation. One limitation of this classification should be obvious to anyone familiar with watershed monitoring: a six-month "pre-ESD" period is too short for establishing statistical relationships between hydrologic variables measured in the impact and control watersheds as has been done elsewhere (e.g., Hornbeck et al. 1970). Further, the six-month period effectively included data collected only in the spring and summer seasons—excluding fall and winter observations. Hydrologists have long recommended multi-year pre-treatment "calibration" periods (e.g., Hewlett et al. 1969), but this is particularly problematic in urban hydrology research where researchers lack any control over the timing and pace of development. This was the case here where the collection of even one complete water year of pre-treatment monitoring data turned out to be infeasible.

Regardless, we hypothesized that the UTLP watershed would show *increases* in baseflow and "old water" runoff, but *reductions* in total runoff, storm event runoff, peak storm event runoff, event runoff ratios, overland flow, "new water" runoff, and event mean concentrations (EMCs) of nitrogen (N) and phosphorus (P). Comparable "paired" data from the PLBR watershed with stable land use would be used to provide hydroclimatological control—thus allowing for the possible application of appropriate statistical tests (i.e., ANCOVA) that involve comparisons of adjusted means for the two time periods.

Field Methods

The DWLR's provided us with continuous, 5-min stream gage height (*G*) data records for both watersheds from the time of station installation (November 2019 at PLBR; March 2020 at UTLP) through September 2023; data were downloaded frequently (at least monthly) to a field laptop computer. With few exceptions, these records were highly accurate and essentially 100% complete. At both sites, a small number of records had to be corrected for a computer date/time error. At UTLP, several short periods of data had to be corrected (by linear interpolation) when the float on the DWLR became hung up in the stilling well during hydrograph recessions; in another case, UTLP data

that had been lost due to a programming error were corrected by substituting back-up water level data from the Aquatroll 500.

"Wading" measurements of instantaneous stream discharge (i.e., the USGS "mid-section method"; Rantz et al. 1982) were made regularly during the project to support the development of rating curves that could be used to convert gage height (*G*, in m) readings to discharge (*Q*, in $m^3 s^{-1}$ or "cms"). For the field measurements, we used: 1) a Marsh-McBirney digital electromagnetic current meter to measure velocity; 2) a wading rod to measure depth and to locate the electromagnetic probe in the vertical dimension (0.6 times the depth from the water surface); and 3) a tape or tag line to measure the width of each subsection. Each complete measurement typically required 15-30 individual subsection measurements depending on the discharge. We made the discharge measurements as frequently as possible and across a full range of stage conditions (i.e., low baseflow through peak stormflow—including during flash flood conditions!), but the combination of time of travel between our lab and the field sites, COVID-19 research restrictions that were in effect near the beginning of the project, and the flashiness of these streams made this a challenging proposition. Nonetheless, we were able to successfully derive solid linear rating curves (i.e., *logQ* vs. *log(G-e)* relationships) from the field data using established methods (Rantz et al., 1982); offset parameters (i.e., *e*) were obtained by trial and error during the linearization process. Importantly, application of the curves to the field data required minimal extrapolation.

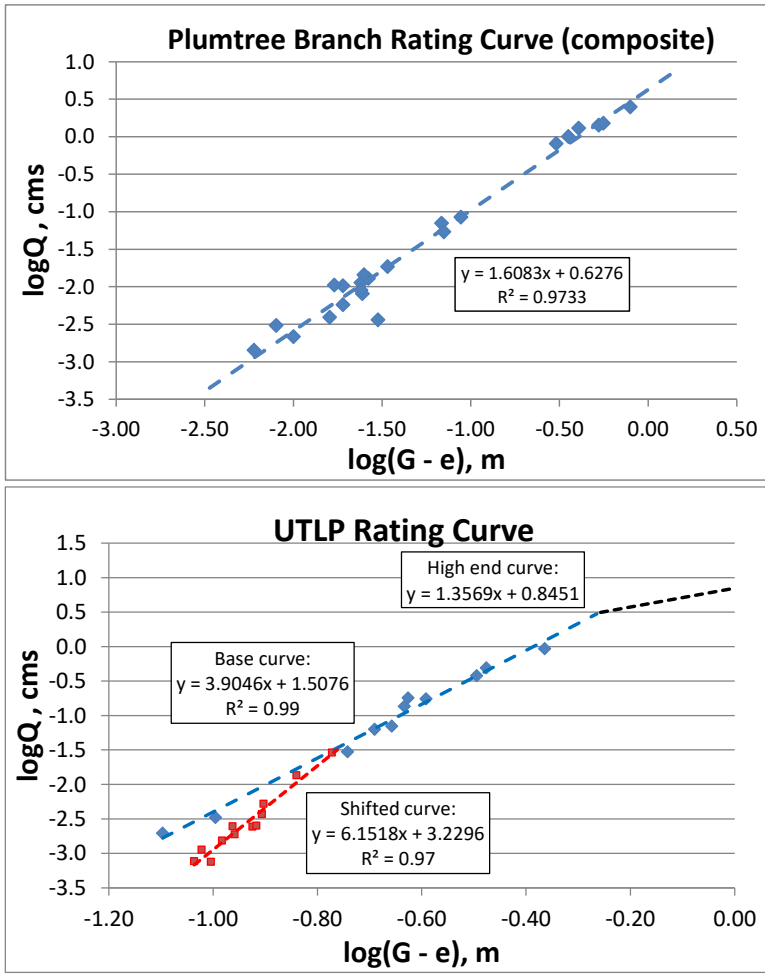

Figure 6. Rating curves for the PLBR (top panel) and UTLP (bottom panel) gaging stations.

For example, we calculated that the highest actual discharge measurements made at PLBR and UTLP were exceeded less than 0.017% (~6 hr) and 0.034% (~11 hr), respectively, of the total time that these sites were recording data during the 3.5-year project. In the case of PLBR, several channel scour and sedimentation events necessitated adjustment of the offset (*e*) from the base rating curve to accommodate these shifts. At UTLP, a flash flood event in August 2020 caused a major shape change at the station control—which we reconstructed in the aftermath of the event—and the addition of a shifted section of the rating curve which we were able to establish with subsequent field discharge measurements. We also added a "high end curve" to reflect the influence of channel control under bankfull conditions (Figure 6).

Unfortunately, despite regular (~ bi-monthly) calibration, the Aquatroll 500s (ATs) proved far less reliable than the DWLRs, particularly for producing accurate turbidity data. At both stations the ATs had to be pulled for service about midway through the project, costing us several months of data. The conductivity and turbidity probes on both units were each replaced twice during the course of the project when they stopped producing data. Even more problematic was repetitive biofouling, particularly at the PLBR station, that necessitated very frequent cleaning and recalibrating of the turbidity probe. For these reasons, we religiously cleaned the ATs on each site visit (in particular, at the beginning and end of each storm event that was sampled). As part of data validation, we compared *in situ* SC data against lab SC values for samples collected contemporaneously and found very good agreement at both sites (figure not shown). At the UTLP station, bedload transport of sand and fine gravel often caused the AT there to become "sedimented in" under high flow conditions; the sedimentation adversely affected both the turbidity and conductivity probes (but not the pressure or temperature data). Finally, the tubes that were used for powering the ATs and for telemetering data failed several times as well—costing additional data.

The tipping bucket rain gages located at "Kappa" and "Delta" stations were nearly 100% reliable in producing 10-min digital rainfall data over the course of the project. The only exception were short periods of snowfall (where some data were likely lost) and a situation where one of the funnels became clogged by insect debris). Since this project focused on rainfall (not snowmelt) events and Ellicott City received very little frozen precipitation during the project, the first issue was not that important. Where data were lost at one station, we were able to substitute data from the other

station. Moreover, for purposes of interpreting storm event data, we made the decision very early on in the project to estimate "areal" event rainfall for the watersheds rather than relying solely on the tipping bucket "point" rainfall data. This decision was forced by the occurrence of several events during the summers of 2020 and 2021 in which it was very apparent that one of the watersheds received significantly more (as much as nearly two times) rainfall than the other (despite the fact that they are contiguous). Since we were focused on "pairing" event data from the two watersheds, any differences in areal event rainfall would add an additional error in developing statistical relationships.

We thus estimated areal rainfall for runoff-producing storm events by downloading WSR-88D ("NEXRAD") Level III 5-min rainfall data (instantaneous precipitation rate dual polarization product) generated by the Sterling, Virginia station from either the NCDC or UCAR THREDDS server using NOAA's Weather and Climate Toolkit (WCT)[1]. We used the WCT "point subset tool" to download daily rainfall data for 1) the two point rainfall stations in Ellicott City; 2) ten "pseudo stations" in the PLBR watershed; and 3) six "pseudo stations" in the UTLP watershed. For each station or pseudo station, we integrated the storm rainfall intensity data to produce an (unadjusted) storm event rainfall total. We then used the actual station data to estimate a mean field bias factor and adjusted the pseudo station totals accordingly. Finally, we used the Thiessen polygon method (Chow et al. 1988) to compute areal event rainfall from the gage-adjusted pseudo station values.

Baseflow and stormflow sampling were used to characterize water quality during the study (and supplement the high frequency *in situ* data from the ATs). "Grab" baseflow samples were collected approximately monthly in 1L poly bottles (at least 40 hr following significant precipitation); stormflow samples were automatically collected during selected stormflow events (~ 1 event per month) using the Sigma samplers. We targeted events with forecasted rainfall of 0.5" or greater. Depending on the forecasted duration of a selected event, the samplers were programmed to collect 1L samples at frequencies between 0.5 and 1.0 hr^{-1}; the program allowed us to adequately characterize both rising and falling limbs of each event while typically providing 20-24 samples for laboratory analysis. Samples were placed on ice in coolers and transported to our water chemistry laboratory (WCL) in

[1] The UCAR THREDDS server proved far more reliable than the NOAA server for reasons that we were never really able to resolve. While the NOAA server supposedly archives NEXRAD data permanently, the UCAR server only maintains data for one month. Unfortunately, by the time we realized this, data on the UCAR server for some critical events had been deleted and the data were apparently not archived properly on the NOAA server.

Frostburg, Maryland for subsequent processing, filtering, aliquoting, preservation, analysis, and storage (either refrigeration at 4 deg C or freezing at -10 deg C).

Table 2. Measured water quality constituents and associated laboratory methods, detection limits, and holding times; [1]Flow Injection Analysis System; [2]Ion Chromatography System.

Constituent (units)	Method	Instrument	Detection Limit	Holding Time (days)
Nitrite-N (mg/L)	USEPA (1993) 354.1	Lachat QuikChem Automated FIAS[1]	0.0022	28 (frozen)
Nitrate-N (mg/L)	USEPA (1993) 353.1	Lachat QuikChem Automated FIAS[1]	0.0041	28 (frozen)
Ammonium-N (mg/L)	Fishman (1993)	Lachat QuikChem Automated FIAS[1]	0.0026	28 (frozen)
Total N (TN, mg/L)	APHA (2017) 4500-P (J)	Lachat QuikChem Automated FIAS[1]	0.0216	28 (frozen)
Total dissolved N (TDN, mg/L)	APHA (2017) 4500-P (J)	Lachat QuikChem Automated FIAS[1]	0.0216	28 (frozen)
Particulate-N (mg/L)	APHA (2017) 4500-P (J); by difference (TN – TDN)	Lachat QuikChem Automated FIAS[1]	0.0094	28 (frozen)
Dissolved organic N (DON, mg/L)	APHA (2017) 4500-P (J); by difference (TDN – Nitrate-N – Nitrite-N – Ammonium-N)	Lachat QuikChem Automated FIAS[1]	0.026	28 (frozen)
Orthophosphate-P (mg/L)	USEPA (1993) 365.1	Lachat QuikChem Automated FIAS[1]	0.0031	28 (frozen)
Total P (TP, mg/L)	APHA (2017) 4500-P (J)	Lachat QuikChem Automated FIAS[1]	0.0044	28 (frozen)
Total dissolved P (mg/L)	APHA (2017) 4500-P (J)	Lachat QuikChem Automated FIAS[1]	0.0044	28 (frozen)
Particulate-P (mg/L)	APHA (2017) 4500-P (J); by difference (TP – TDP)	Lachat QuikChem Automated FIAS[1]	0.0077	28 (frozen)
Soluble organic/colloidal-P (SOC-P, mg/L)	APHA (2017) 4500-P (J); by difference (TDP – Orthophosphate-P)	Lachat QuikChem Automated FIAS[1]	0.0053	28 (frozen)
Chloride (mg/L)	USEPA (1993) 300.1	ThermoScientific Dionex Aquaion ICS[2]	0.0032	14 (4 deg C)
Bromide (mg/L)	USEPA (1993) 300.1	ThermoScientific Dionex Aquaion ICS[2]	0.0015	14 (4 deg C)
Nitrate (mg/L as N)	USEPA (1993) 300.1	ThermoScientific Dionex Aquaion ICS[2]	0.0015	14 (4 deg C)
Sulfate (mg/L)	USEPA (1993) 300.1	ThermoScientific Dionex Aquaion ICS[2]	0.002	14 (4 deg C)
Specific conductance, 25°C (SC, µS)	USEPA (1993) 300.1	YSI Conductance Meter w/Cell	0.1	7 (4 deg C)
Total suspended solids (TSS, mg/L)	USEPA (1993) 160.2	Gravimetric	0.3	7 (4 deg C)

Laboratory Methods

Aliquots of collected samples were analyzed for a suite of 18 water quality constituents using established analytical methods within pre-established holding times. Several constituents that lack direct measurement methods (e.g., particulate-N, particulate-P, DON, and SOC-P) were computed by difference from other direct measurement methods (e.g., particulate-N = TN – TDN; particulate-P = TP – TDP). Our laboratory used an extensive array of analytical quality control checks (i.e., analysis of blanks, independent quality control samples, spikes, duplicates, etc.) and other procedures (e.g., error checking) to aid in assuring that the water quality data generated by the project were accurate, precise, and reproducible. Laboratory analytical methods, estimated method detection limits (from statistical analysis of blanks), and holding times for each the water quality constituents measured during the project are shown in Table 2.

Data Management and Analysis

Digital hydrologic (rainfall and runoff) and water quality data from each watershed were uploaded to separate MS-Excel workbooks where the data could be readily appended to existing data records and maintained over the life of the project. After new records were added to a current workbook, we modified the name of the workbook by including the date of the last observation and saved the updated version. Out-of-date workbooks were archived. We used MS-Excel primarily to: 1) generate the rating curves; 2) generate continuous records of both instantaneous and mean hourly discharge—

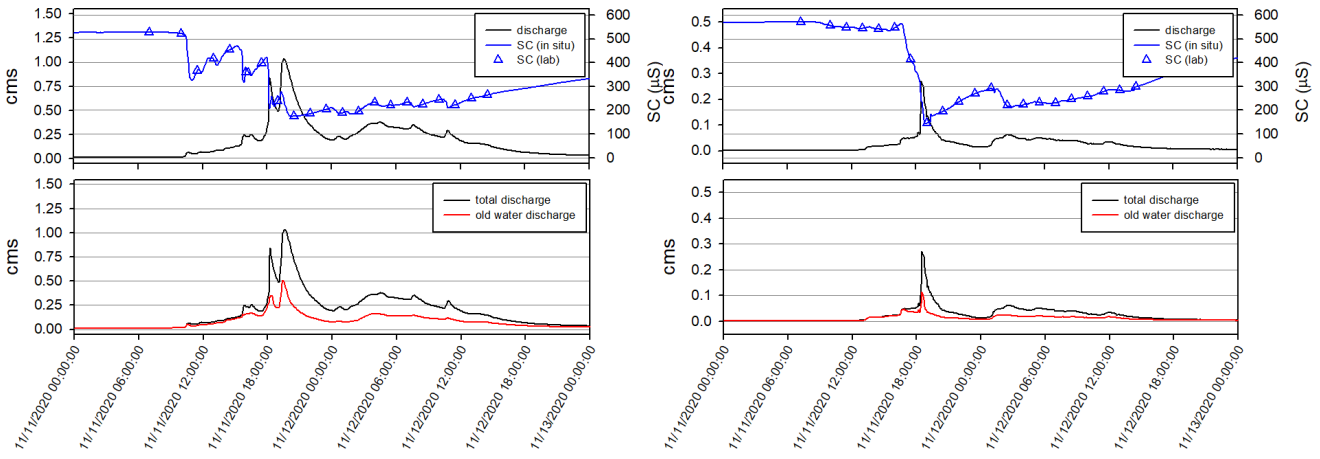

Figure 7. Measured SC (in situ and lab values) and stream discharge (top) and chemical hydrograph separations (bottom) for event "K" (11/11 – 11/12/2020): PLBR (left panels); UTLP (right panels).

the latter used to generate complete records of mean daily discharge that were used to compute annual runoff (water year basis: October 1 – September 30); 3) generate rainfall and runoff statistics for all (n = 101) runoff-producing events; 4) compute daily, monthly, and annual "point" rainfall records; and 5) display the data to facilitate visual error checking. Water quality data generated by the WCL, as well as data downloaded from HydroVu, were uploaded to separate sheets in the workbooks; the water quality data were merged with the runoff data to facilitate computation of EMCs for each of the characterized events (i.e., the lettered events listed in Table A2).

Linked sheets were used to perform chemical hydrograph separations based on variations in specific conductance (SC) measured *in situ* in the two streams (Sklash et al. 1976, Pellerin et al. 2008). In all except a handful of winter events where the effects of road-salting were clearly evident (i.e., SC increased dramatically during these events), SC decreased during storm events and reached minimum values near the time of peak discharge. As discharge receded, SC gradually returned to antecedent values (e.g., event "K" results shown in Figure 7). We used the 5-min SC data from the ATs to separate hydrographs (n = 81 PLBR events, 68 UTLP events) into "new water" and "old water" contributions, using the antecedent baseflow SC as the old water end-member and assuming that the new water SC was zero. The chemical hydrograph separation method assumes that both end-member concentrations are constant during each event. New water and old water runoff hydrographs were generated for each event and the hydrographs were numerically time-integrated and normalized to compute new water and old water runoff depths (m). For each event or which chemical hydrograph separation could be performed, we used the method of Eshleman et al. (1993) to compute the new water contributing area (NWCA).

We downloaded mean daily discharge data for several long-term USGS gaging stations located in the vicinity of the ULPRB to enable comparisons of annual runoff depths (mm yr^{-1}) for the PLBR and UTLP watersheds (and provide a validation of our stream gaging effort). NOAA monthly rainfall data for the station at BWI airport (www.weather.gov as an autogen file bwiprecip.pdf) were also downloaded and used to estimate deviation from long-term climatological normals.

In addition, we used the LOADEST program developed by USGS (Runkel et al. 2004; 2013 update with goodness-of-fit statistics) to estimate monthly and annual loads of each water quality constituent for both watersheds. We developed seven-parameter models based on the technique of adjusted

maximum likelihood estimation (AMLE) using calibration files that included the complete record of laboratory water quality analyses coupled with the estimated instantaneous discharge at the time of sample collection. For some of the constituents, it was necessary to censor the water quality data to reflect the fact that some observations were below the method detection limit. This is not a significant problem for LOADEST, but it does effectively reduce the sample size on which the associated model is based. Estimation files were created from our complete instantaneous hourly discharge datasets. Monthly and annualized loads (and discharge-weighted concentrations) were computed from the LOADEST output and expressed on a per hectare basis (i.e., in kg/ha-yr).

We also performed continuous hydrograph separations for both watersheds using the recursive digital filter (RDF) method of Eckhardt (2005); *a* parameter values were obtained from recession analysis of streamflow hydrographs for long (i.e., multi-week) periods without measurable precipitation. BFI_{max} parameters were estimated through an optimization process using new water contributions based on the chemical hydrograph separations as proposed by Zhang et al. (2013) and Lott and Stewart (2016). Our method is similar to the approach demonstrated by Foks et al. (2019), although our approach is based on literally tens of thousands of SC observations from each watershed (and could thus be expected to produce more accurate results). We used the optimized parameter values and the RDF method (https://sephydro.hydrotools.tech/pageMain.php) to produce continuous daily estimates of direct runoff (DR) and baseflow (BF) for both watersheds; model output was annualized (water year basis) for comparative purposes.

Simple statistical tests and hypothesis testing (e.g., paired t-tests, etc.) were mostly performed using operations available in MS-Excel (Data Analysis add-in); ANCOVA and non-parametric statistical tests were performed using an on-line tool (http://vassarstats.net/).

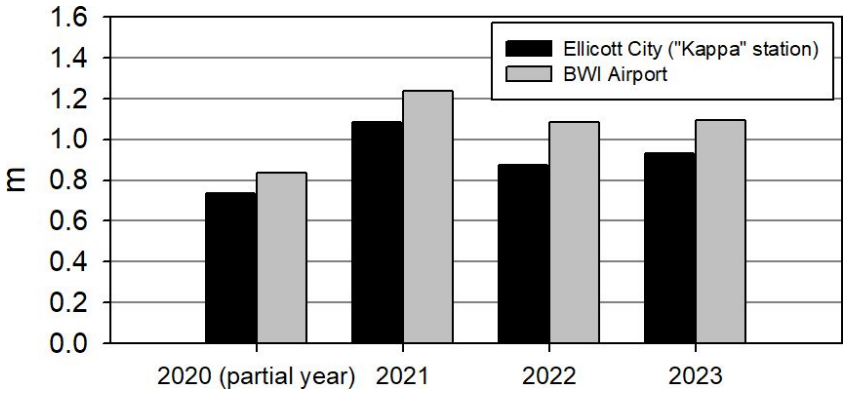

Figure 8. Annual (water year) precipitation for Ellicott City ("Kappa" station) and BWI Airport during the study.

Results

Precipitation. Annual gage precipitation computed for the "Kappa" station in Ellicott City for the 3.5-year project period illustrated significant year-to-year variability that mirrored comparable results reported for the BWI Airport station. Annual precipitation at BWI Airport was consistently (10 – 25%) higher than at our Ellicott City ("Kappa") station—reflecting 1) possible under-catch of liquid or solid precipitation or both; or 2) a "real" decline in precipitation with distance away from the influence of Chesapeake Bay as has been documented by Smith et al. (2012). Using the long-term monthly normals for BWI Airport based on 1981-2010 data, precipitation at BWI was 0.837m (136% of normal) in the partial 2020 water year, 1.237m (108% of normal) in water year 2021, 1.084m (95% of normal) in water year 2022, and 1.09m (96% of normal) in water year 2023. Precipitation at the Ellicott City station was 0.737m, 1.086m, 0.876m, and 0.930m, respectively, for the same periods. If it is assumed that there is no difference in long-term normal precipitation between the two stations, however, then the precipitation at Ellicott City was 120%, 95%, 77%, and 81% of normal, respectively, for the same time periods. If this is correct, then only the partial 2020 water year (that happens to be the entire pre-ESD period) received precipitation that was significantly greater than normal (Figure 8).

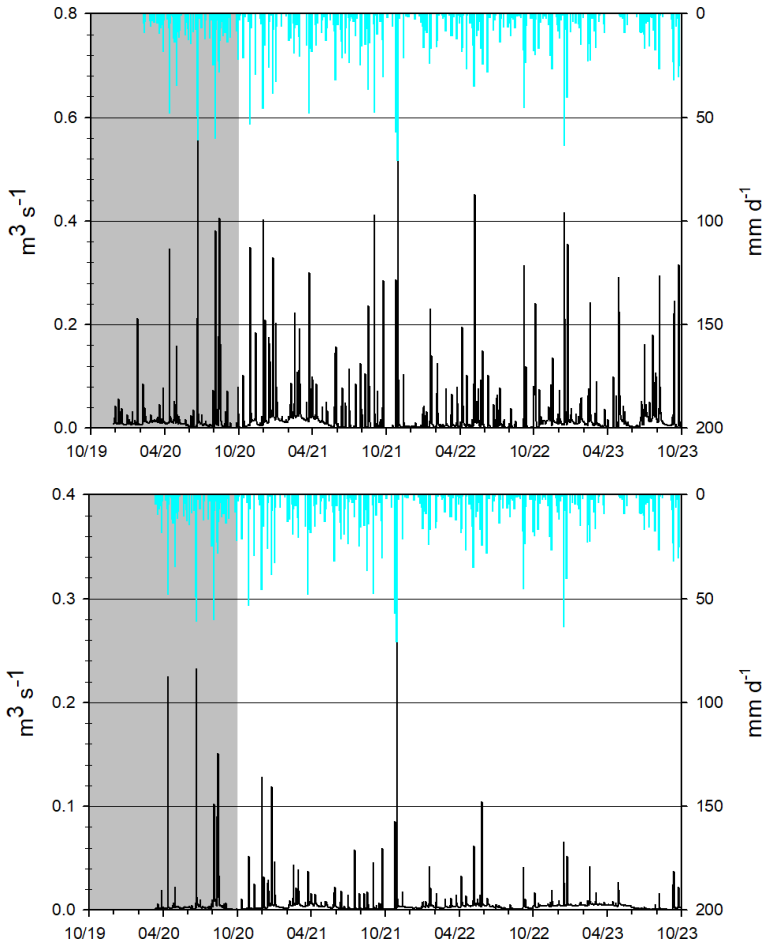

Figure 9. Hyetographs ("Kappa" station) and hydrographs for the entire period of record for PLBR (top panel) and UTLP (bottom panel). The grey shaded period prior to October 2020 is considered the pre-ESD period of the study.

Annual runoff. Daily precipitation ("Kappa" station) and mean daily discharge for both watersheds for the entire 3.5-year study period are shown in Figure 9. Visually, the PLBR hydrograph suggests stationary hydrologic behavior consistent with stable land use, while the UTLP hydrograph suggests a fairly dramatic decline in stormflow runoff and peak discharges. The flood of record at UTLP occurred in the middle of the record (10/29/2021) and only one subsequent storm (5/27/2022) produced a mean daily discharge that exceeded 0.1 cms. Prior to the October 2021 flood, however, six separate storms produced mean daily discharges exceeding 0.1 cms; interestingly, four of these storms occurred in the pre-ESD period; Figure 9).

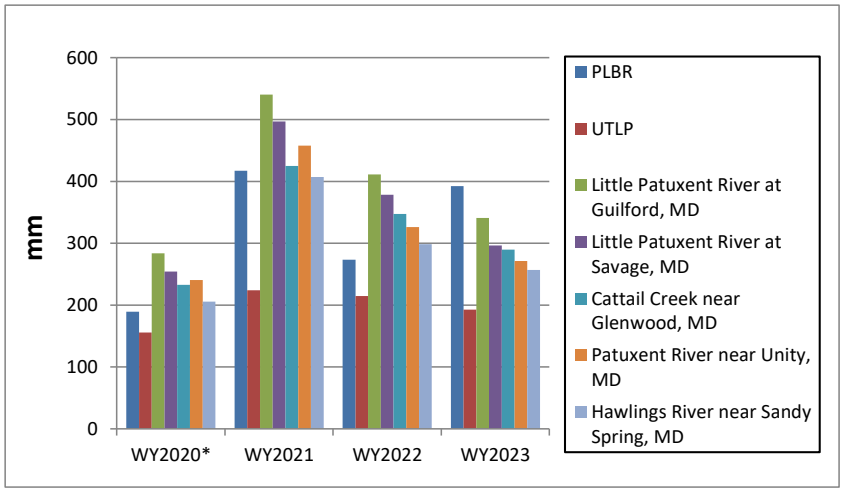

Figure 10. Annualized runoff for water years 2020, 2021, 2022, and 2023 for PLBR, UTLP, and five nearby watersheds gaged by USGS. *Results for water year 2020 are for the six-month period from April through September.

Annual runoff (normalized by area) from the PLBR watershed showed considerable year-to-year variability over the study—likely due primarily to interannual variability in precipitation. The inter-annual runoff variability was similar to the variability observed in data from five watersheds gaged by USGS (based on mean daily discharge data downloaded from the NWIS website (https://waterdata.usgs.gov/md/nwis). For water years 2020, 2021, and 2022, PLBR runoff was consistently at the low end of runoff observed at the USGS watersheds. In 2022, PLBR was considerably higher than at the USGS stations, however, likely reflecting spatial variability in precipitation. Comparing the aggregated runoff for the entire 3.5-year study period, however, PLBR runoff (1,272mm) fell in the middle of the range based on the USGS watersheds; the maximum aggregated runoff was computed for the Little Patuxent River at Guilford, Maryland station (1,577mm), while the minimum value was computed for the Hawlings River near Sandy Spring, MD station (1,169mm). The aggregated PLBR runoff was actually within 2% of the runoff computed for the Cattail Creek near Glenwood, Maryland (1,295mm) and the Patuxent River near Unity, Maryland

(1,296mm) stations. Our interpretation is that these comparisons essentially validate the overall stream gaging effort at PLBR (Figure 10).

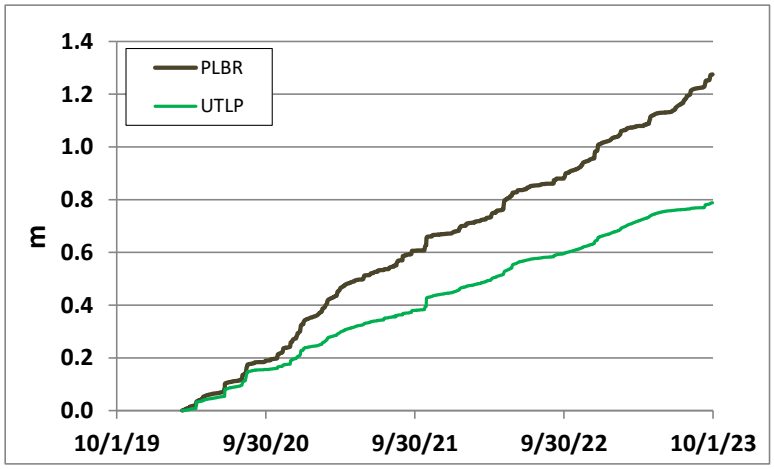

Figure 11. Cumulative runoff from the two watersheds over the life of the project.

Annual runoff from the UTLP watershed was consistently less than at PLBR and the USGS watersheds for the study years. In water year 2020, UTLP runoff (156mm) was about 17% lower than at PLBR (189mm), but the difference between the two watersheds dramatically widened in the full water year 2021 (224mm at UTLP vs. 418mm at PLBR)—a nearly 50% difference. The difference (215mm at UTLP vs. 286mm at PLBR) narrowed slightly in water year 2022—a difference of about 25%. The difference widened again in water year 2023 (193mm at UTLP vs. 392mm at PLBR)—a difference of slightly more than 50% (Figure 10). Over the entire period of ESD implementation, the average difference between the two watersheds was about 42%. We suggest that the runoff difference between these watersheds is a runoff anomaly—not an artifact of our analytical methodology, nor a reflection of a dramatic difference in precipitation regime (although we were able to demonstrate using NEXRAD data that at least one major rainfall event (June 20-21, 2020) produced considerably less areal rainfall at UTLP (2.31cm) than at PLBR (3.35cm; Table A2).

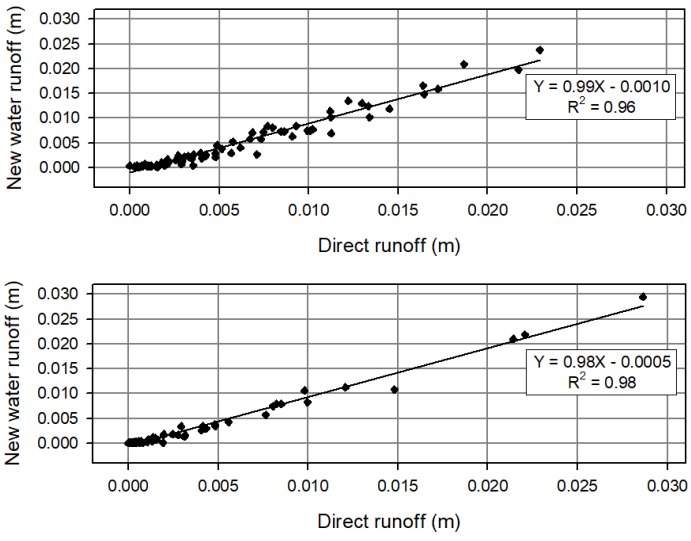

Figure 12. New water runoff from the SC hydrograph separations vs. direct runoff from the digital recursive filter (DRF) after optimization of BFI_{max} parameters for the two watersheds: PLBR (top panel), UTLP (bottom panel). The optimized RDF parameters were: PLBR (a = 0.886; BFI_{max} = 0.675); UTLP (a = 0.957; BFI_{max} = 0.750).

We characterized the anomaly as the cumulative difference (i.e., "running")

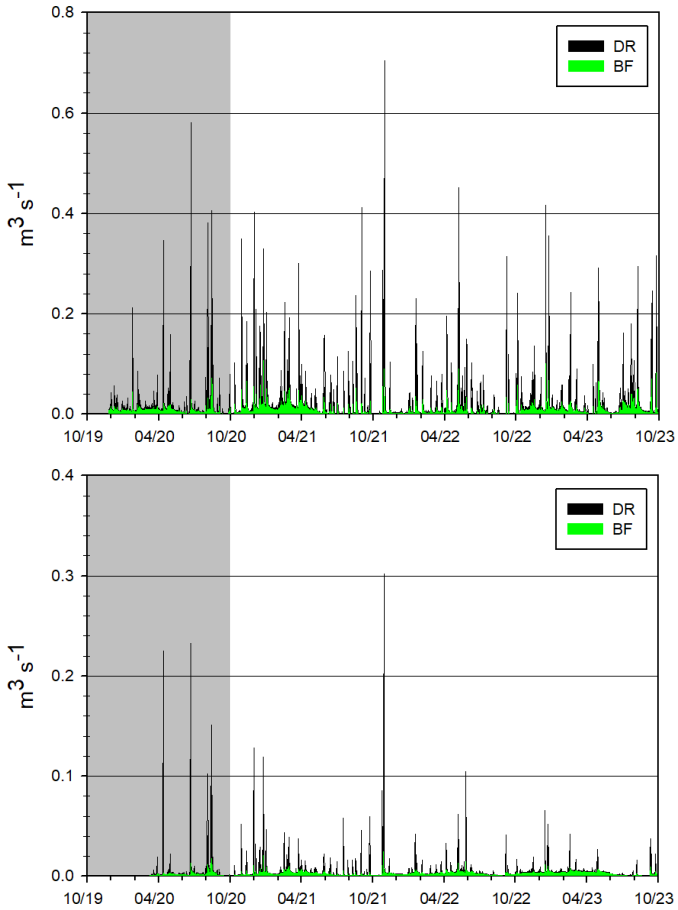

Figure 13. Direct runoff (DR) and baseflow runoff (BF) hydrographs for the two watersheds: PLBR (top panel), UTLP (bottom panel).

difference in daily runoff between the two watersheds over the total project period. As of October 1, 2023, the cumulative difference was 486mm. While the results seem to confirm that such an anomaly exists (Figure 11), the graph does not appear by itself to enable identification of a date at which the apparent anomaly began. While there may be other explanations for the observed runoff anomaly, the results provide support for the hypothesis that ESD implementation can contribute to a decline in total runoff.

Hydrograph separation and baseflow index. After optimization of the BFI_{max} parameter values, we observed very strong statistically significant linear relationships between direct runoff (DR) based on the RDF separation and new water runoff based on the SC hydrograph separation technique for both watersheds; in both cases the slopes of the linear regressions were nearly unity and intercepts were very close to zero (Figure 12). It should be noted that the *a* and BFI_{max} parameter values are both within expected ranges; in fact, the UTLP value of BFI_{max} (0.750) is the suggested default value based on the pioneering work of Eckhardt (2005). The parameterized RDFs were then used to generate DR and baseflow (BF) hydrographs (mean daily discharge) for both streams for the entire study period. As was the case for the total hydrographs (Figure 9), the separated PLBR hydrographs visually suggest approximate stationarity, while the separated UTLP hydrographs suggest non-stationary behavior (Figure 13). In particular, the UTLP BF hydrograph suggests increasing dominance of BF relative to DR over the 3.5-year period. Our next step was to formally test whether this was the case by time-integrating both hydrographs for each watershed and computing annual values of DR, BF, and the baseflow index (BFI = the ratio of annual baseflow runoff to total runoff; BF/R_A).

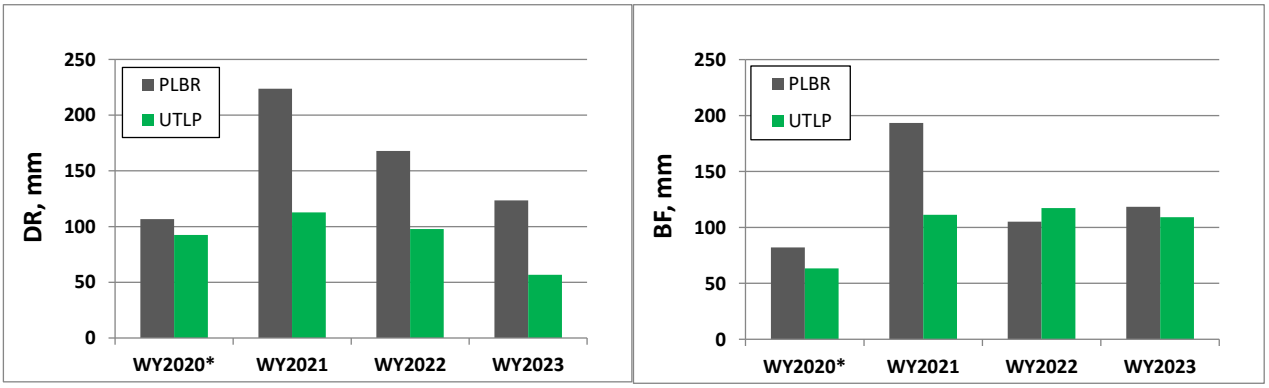

Figure 14. Annual direct runoff (DR; left panel) and baseflow runoff (BF; right panel) for the two watersheds based on optimized two-parameter RDF's; *WY2020 is a partial year (data from March 9 – September 30, 2020).

The computed annual DR from UTLP in partial year 2020 (the pre-ESD period) was only about 13% less than at PLBR, but this gap widened dramatically in subsequent water years (i.e., during ESD implementation): the differences were 50%, 42%, and 54% in water years 2021, 2022, and 2023, respectively—consistent with the apparent diminishment of stormflow runoff suggested by Figure 13. Annual BF showed a different pattern: in 2020, BF from UTLP was about 23% less than at PLBR; as with DR, the BF gap between the two watersheds widened appreciably in WY2021 with BF showing a difference of about 42%. In water year 2022, however, BF was actually about 11% higher at UTLP than at PLBR; in water year 2023, UTLP BF was about 8% lower than at PLBR (Figure 14).

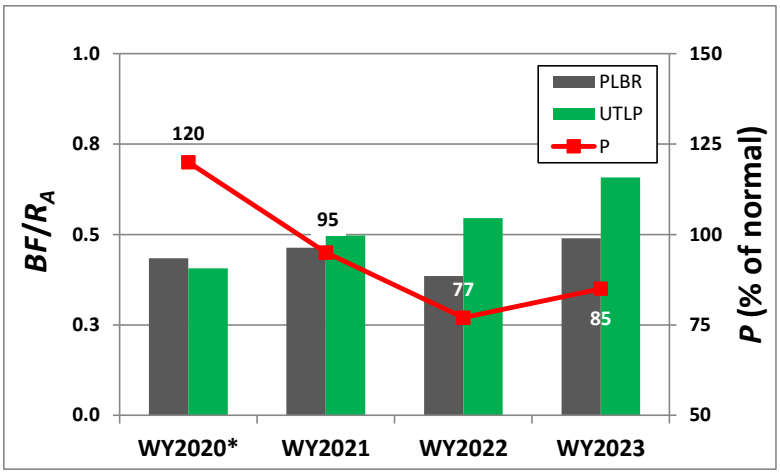

Figure 15. Annual values of the baseflow index (BFI = BF/R_A) for the two watersheds based on application of the optimized two-parameter RDF's. Graph of annual precipitation (Kappa station) expressed as a percentage of normal is also shown. *WY2020 is a partial year (data from March 9 – September 30, 2020).

A better was of examining differences in these runoff components is by comparing the baseflow index (BFI = BF/RA) during the two time periods. In water year 2020, BFI values estimated for UTLP and PLBR were 0.41 and 0.44, respectively (a difference of only about 7%). In the subsequent full water years, the BFI values for UTLP showed a monotonic increasing trend—to 0.50, 0.55, and 0.66, respectively. At PLBR, however,

the range of BFI was much tighter (0.39 to 0.49) with no trend evident in the data—a result that is also seemingly consistent with the stable land use in the control watershed. It is also interesting that increasing BFI at UTLP appears to be largely independent of hydroclimatic conditions as represented by annual precipitation variations (Figure 15).

Storm event hydrologic response metrics. We characterized 101 stormflow-producing rainfall events during the 3.5-year study (see Table A2). We used the extant data to compute a variety of useful metrics for characterizing the hydrologic responsiveness of the two watersheds to individual rainfall events. These metrics are: 1) total storm event runoff (i.e., "volume"); 2) event runoff ratio; 3) peak event runoff; 4) maximum new water runoff; and 5) new water contributing area (NWCA). Total event runoff, event runoff ratio, and peak event runoff were all based on the (total) discharge measured at the watershed outlets (normalized by the respective watershed areas), while maximum new water runoff and NWCA metrics were based on computed measures of new water discharge from the two-component hydrochemical separations using SC as the "conservative" tracer as others have proposed for urban watersheds. The NWCA represents the volume of new water discharged during an event, normalized by the event rainfall depth and expressed as a percentage of the watershed area (Eshleman et al. 1993). Since events A, B, and C were not characterized at UTLP, the sample size of events that could be "paired" (i.e., data for both sites were available) was reduced from 101 to 98. As noted previously, event runoff ratio metrics were based on gage-adjusted areal rainfall values that incorporated a NEXRAD Level III precipitation product; this allowed us with the possibility of eliminating those storm events that showed large differences in areal rainfall between the two watersheds prior to performing any paired statistical analysis. Those events (n = 12) with CV of areal rainfall that exceeded 0.25 were considered outliers for the paired analyses only. We substituted Kappa rainfall data for those events for which NEXRAD Level III data could not be obtained (n = 5), effectively assuming rainfall homogeneity. In the end, we used 86 common events (14 pre-ESD, 72 during-ESD) for the pair-wise analyses that did not require SC separation data. Maximum new water runoff and NWCA metrics could only be computed for a smaller subset of events, however. This is due to the fact that many winter events did not show SC dilution as discussed earlier (i.e., these events violated the assumptions of the hydrochemical separation method); another subset of events lacked a complete record of *in situ* SC data at one of the two stations. Overall, fewer than 70 of the 86 events had sufficient hydrochemical data to allow for

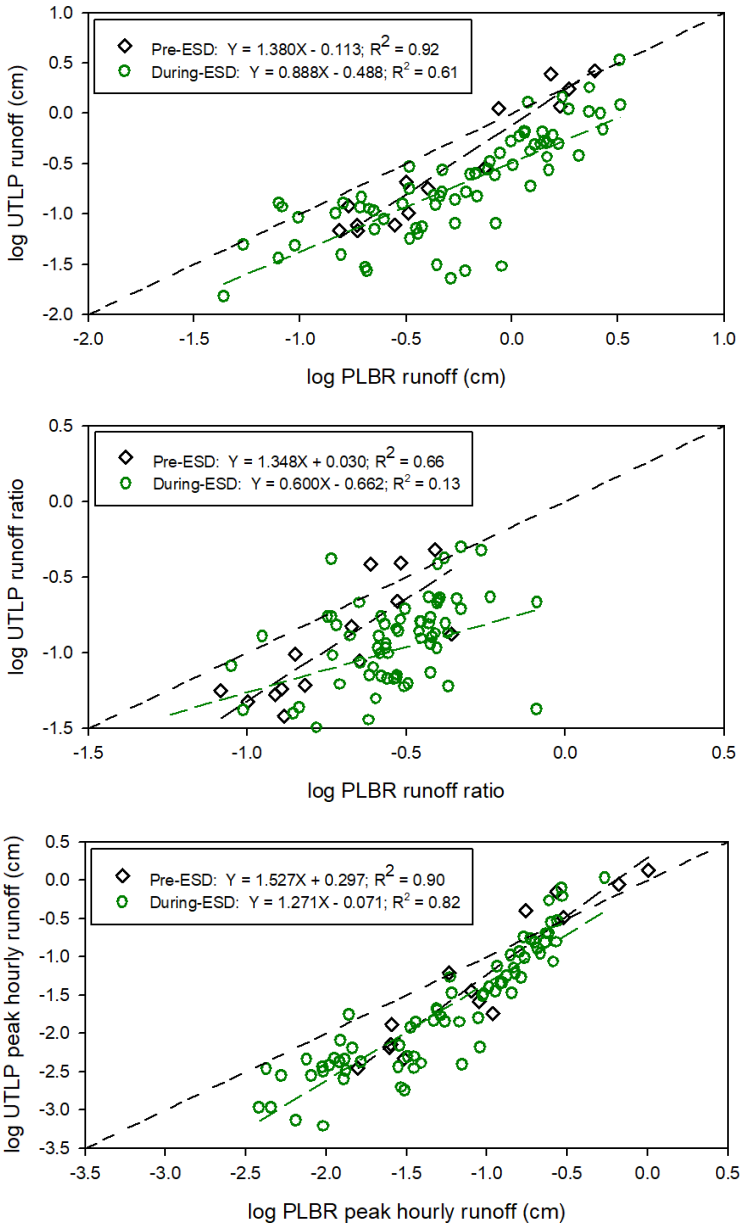

Figure 16. Paired comparison of (log base 10) total storm event runoff (upper panel); storm event runoff ratio (middle panel); and peak hourly storm runoff (bottom panel) for the group of 86 common stormflow-producing rainfall events; linear regression statistics are shown in the legends.

computation of these two metrics for both stations; for purposes of pairing events, only 58 of these were common events, however.

The group of 86 common events included two relatively infrequent events: 1) a 6.1cm rainfall event on 6/22-23/2020 with a 20-year recurrence interval for one-hour rainfall (5.97cm); and 2) a rainstorm on 8/3–4/2020 associated with Tropical Storm Isaias that produced 8.0 cm of rainfall over 24 hours (Table A2). Archived NEXRAD data were unavailable for the latter event so we assumed homogenous rainfall between the watersheds. A 5cm rainfall event on 6/20-21/2020 with a 5-year recurrence interval for one-hour rainfall (4.62cm) was included in the "outlier" group based on NEXRAD data showing high variability in areal rainfall between the two watersheds. All other events had estimated recurrence intervals of less than one year for both one-hour and 24-hour rainfall according to NOAA Atlas 14 (Bonnin et al. 2004). An interesting coincidence is that all three of these infrequent events occurred during the pre-ESD period.

Table 3. Storm event hydrologic response metrics for the two time periods. Results are median paired differences between the two watersheds based on common events: PLBR – UTLP. No. of events (n) shown in parentheses; U and P values from one-tailed Mann-Whitney U tests.

Response metric	Pre-ESD (n)	During-ESD (n)	U	P
Storm event runoff (cm)	0.11 (14)	0.42 (72)	247	< 0.01
Store event runoff ratio (dimensionless)	0.067 (14)	0.18 (72)	202	< 0.001
Peak hourly storm event runoff (cm)	0.012 (14)	0.027 (72)	342	< 0.05
Maximum new water runoff (%)	-1.6 (8)	14.8 (50)	121	< 0.05
New water contributing area (%)	3.4 (8)	8.0 (50)	98	< 0.05

We attempted to analyze the paired event metrics using ANCOVA—a parametric statistical test that can be considered a hybrid of linear regression and ANOVA. Essentially, where linear regression can identify statistically significant relationships between paired independent observations, ANCOVA can determine whether differences in adjusted means for two groups of observations (e.g., two time periods) are statistically significant. To apply ANCOVA correctly, the data must also pass a homogeneity of regression test. We used P-values of 0.05 to test for the significance of the differences in adjusted means, as well as for the regression homogeneity test. Not surprisingly, we found that the total event runoff, event runoff ratio, and peak event runoff metrics were highly skewed, so we performed the ANCOVA tests on log-transformed data. While statistically significant regressions were obtained, the ANCOVA results for (log) storm event runoff failed the regression homogeneity test. Further, while the ANCOVA results for both (log) storm event runoff ratio and (log) peak hourly storm runoff passed the homogeneity test, the differences

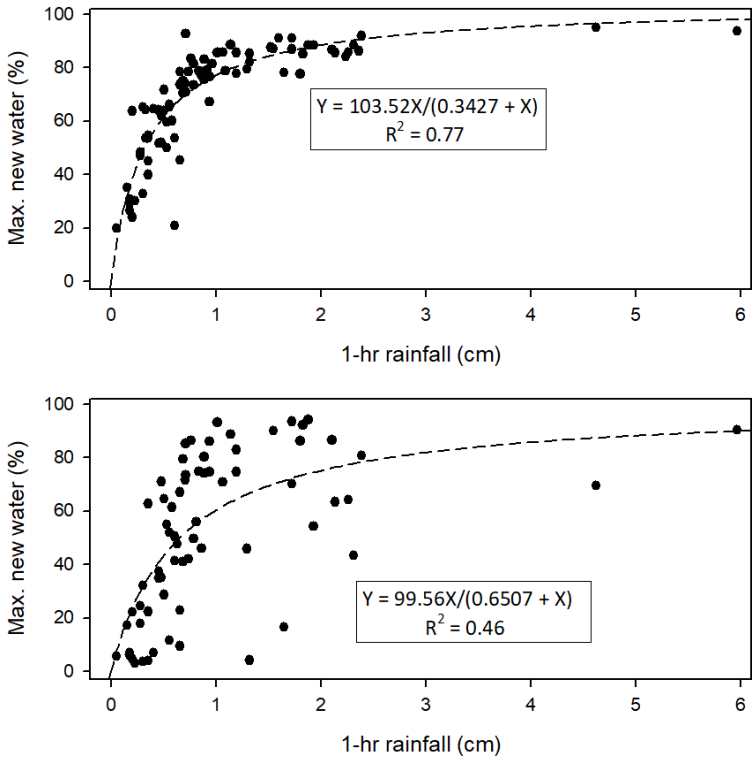

Figure 17. Relationships between maximum new water %) and one-hour rainfall for PLBR (top panel) and UTLP (lower panel). Two-parameter hyperbolas fit to the data using regression are also shown.

between the two time periods were not significantly different (P = 0.07 for runoff ratio; P = 0.34 for hourly runoff; Figure 16).

These results led us to consider a non-parametric test (i.e., Mann-Whitney U test) that would not be as hampered by the small sample size for the pre-ESD period and the regression homogeneity requirement. We tested for changes in median paired differences (PLBR – UTLP) in the same three runoff metrics between the two time periods (one-tailed tests); the data were approximately normally distributed and all three metrics produced statistically significant results. The median difference in storm runoff increased by nearly a factor of four between the two time periods (from 0.11cm to 0.42cm), while the median difference in runoff ratio nearly tripled (from 0.07 to 0.18). The median difference in peak hourly runoff more than doubled (from 0.012cm to 0.027cm). Similarly,

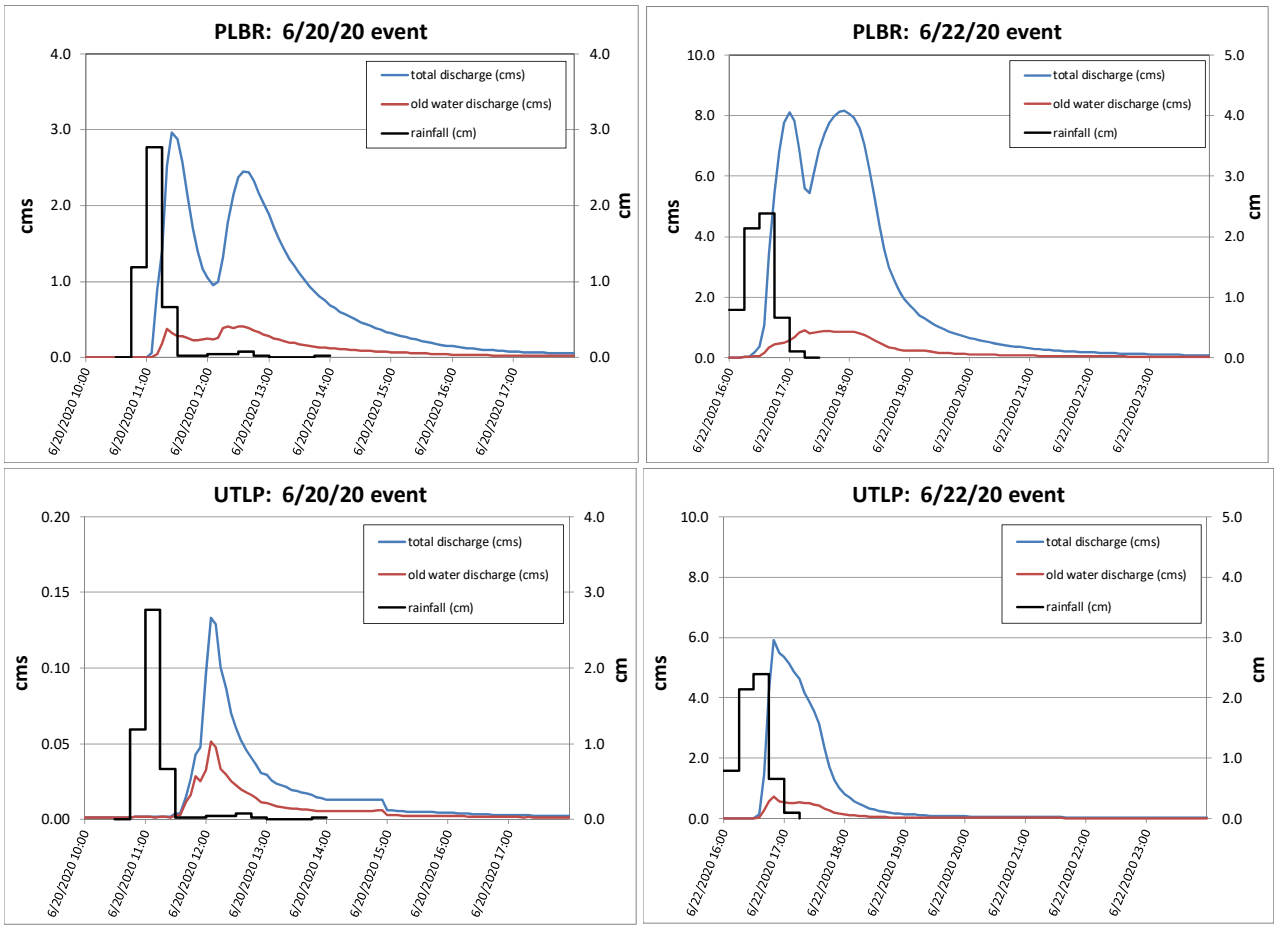

Figure 18. Hydrologic data (rainfall, total discharge, and old water discharge) and hydrochemical separation results for the 6/20-21/20 and 6/22-23/20 events at PLBR (top panels) and UTLP (lower panels). Note differences in y-axis scaling. Similar results obtained for the 7/17/21 event (not shown).

despite even smaller sample sizes (8 pre-ESD events, 50 during-ESD events), we also found statistically significant increases in median differences in maximum new water percentage (-1.6% to 15%) and new water contributing area (NWCA) percentage (3.4% to 8.0%) between the time periods (Table 3). The statistically significant increases in the median differences in each these metrics are consistent with our hypotheses regarding the impact of ESD implementation in the UTLP watershed. We also determined that maximum hourly rainfall (intensity) was the best predictor of maximum new water in both watersheds. Rainfall intensities that exceeded 2.0 cm/hr usually produced event peaks characterized by at least 80% new water; several of the events actually produced maximum new water contributions above 90%. The highest values of the metric asymptotically approached values close to 100 with high one-hour rainfall. Two-parameter hyperbolic functions provided excellent fits to the data from both watersheds, although the PLBR model fit (R^2 = 0.77) was much better than the UTLP model fit (R^2 = 0.46; Figure 17).

Unit hydrographs. We used data from three high-intensity, short duration events (6/20-21/20; 6/22-23/20; 7/17/21) to derive 0.5-hr unit hydrographs for the two watersheds; (coincidentally, two of the events occurred just two days apart!). We made one modification to the standard method by using the hydrochemical separation based on SC to determine the "new water" runoff (i.e., direct runoff) hydrographs (Figure 18). We employed the Ø-index approach (i.e., assuming a constant rate of abstraction, Ø) to identify the specific rainfall pulse(s) that produced direct runoff. Due to the selection of events, we mostly avoided having to compute the unitgraphs by a complex deconvolution process. In fact, for the PLBR watershed for the 6/20-21/20 event and for the UTLP watershed for the 7/17/21 event, we actually determined that only one 0.25-hr pulse contributed direct runoff, so the resultant 0.5-hr unitgraph was computed by simple convolution of

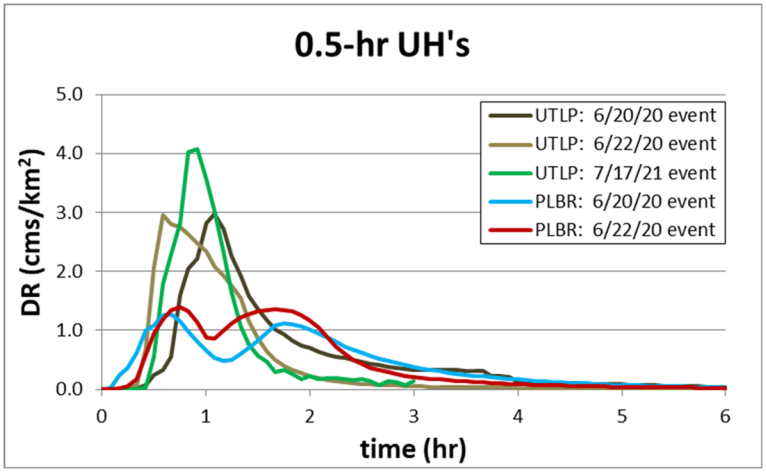

Figure 19. PLBR and UTLP 0.5-hr unit (1.0 cm) hydrographs derived from hydrologic data collected during the project (see Figure 18 for displays of the actual storm data from which the unitgraphs were computed). Note that x- and y-axes are scaled identically.

responses to two consecutive 0.25-hr, 0.5cm excess rainfall pulses. The use of data from multiple events allows us to examine both within- and among-watershed variability in runoff responsiveness to "unit" inputs of excess rainfall.

Visually, the most obvious difference is that the PLBR unit hydrographs are both double-peaked, while the UTLP unit hydrographs are all single-peaked (Figure 19). This difference is significant, because it means that the single-peaked unitgraphs derived for the UTLP watershed are both appreciably higher (i.e., less attenuated) than the unitgraphs derived for PLBR by a factor of two to three. Obviously, if the double peaks observed at PLBR were superimposed, then the differences between the watersheds would be much less significant. It must also be noted that despite the fact that UTLP unit hydrographs were derived both for the pre-ESD and during-ESD periods, we are not comfortable attributing differences in hydrograph shape to ESD implementation.

Water quality. In addition to the hundreds of thousands of records of 5-min stream temperature, SC, and turbidity data, we collected and fully analyzed 633 discrete water samples from PLBR and 557 samples from UTLP during the 3.5-year project including both baseflow and stormflow samples. Overall, 28 common events were characterized, in addition to three events (A, B, and C) that were sampled only at PLBR in late 2019 and early 2020 prior to installation of our monitoring station at UTLP (Table A2). For each of the events, the frequency of sampling and analysis was apparently sufficient to characterize overall water quality variations and support computation of event mean concentrations (EMCs). As in the case of the hydrologic results, the water quality results suffered from the very short pre-ESD implementation period; only five events in the pre-ESD period

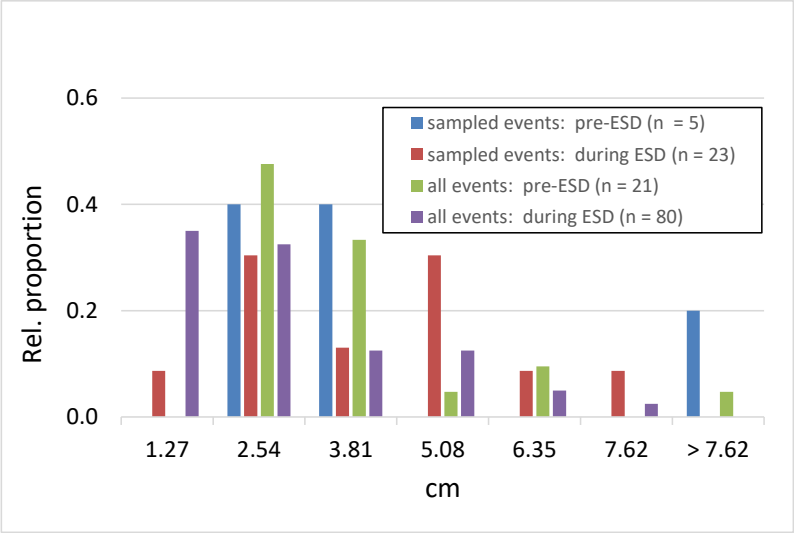

Figure 20. Relative frequency distributions for sampled storms and all storms by time period based on 24-hour rainfall measured at the Kappa station.

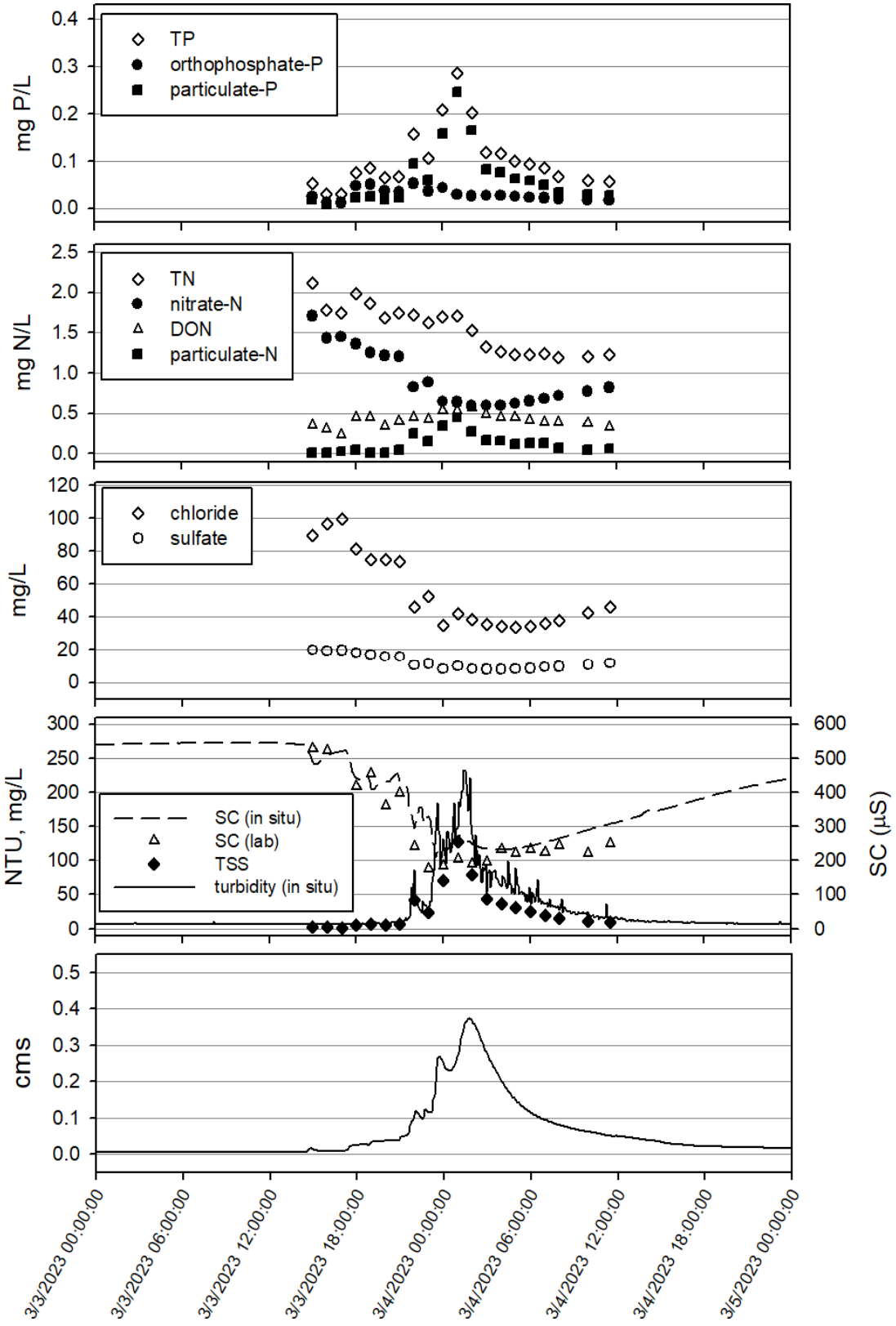

Figure 21. Changes in key water quality constituents during event AE (3/3 – 4/23) at the PLBR station.

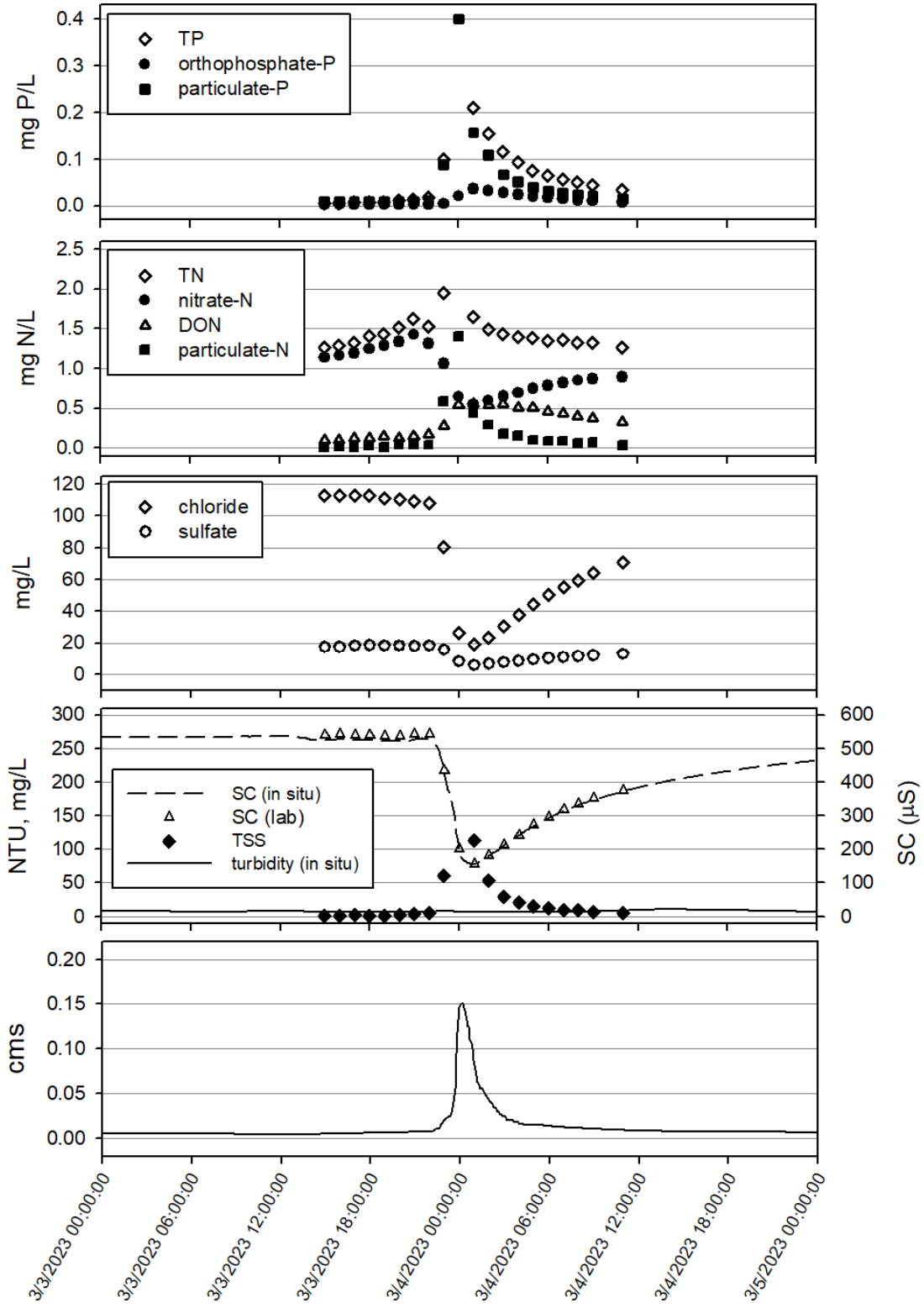

Figure 22. Changes in key water quality constituents during event AE (3/3 – 4/23) at the UTLP station.

were characterized for water quality (although this number represented 24% of all of the stormflow-producing events during the six-month period). The percentage is only slightly less than the 29% of all events that produced stormflow in the during ESD period.

The other issue is related to the types of storm events that occurred in each period. For the hydrologic results based on all common events for both periods, we were able to effectively evaluate differences in stormflow metrics between the two stations and compare the differences among the two time periods. In the case of the water quality data with even smaller sample sizes, however, differences in the types of events actually characterized could potentially play a more important role. It turns out that there were no storms in the pre-ESD period in which 24-hour rainfall ("Kappa" station) was less than 1.27cm (0.5in); obviously, none were sampled for water quality. For the during-ESD period, however, these storms were the most common (35% of all storms) and about 9% of the storms that we sampled were of this type. For storms with 24-hour rainfall between 1.27cm and 2.54cm (1.0in), there was much less variation between periods and we sampled similar percentages: 1) pre-ESD sampled storms comprised 40% of the total number, compared to 48% of all storms; and 2) during-ESD sampled storms comprised 30% of the total number, compared to 33% of all storms. A similar result was obtained for storms with 24-hour rainfall between 2.54cm and 3.81cm (1.5in), although there were fairly large differences between the two periods: 1) pre-ESD sampled storms comprised 40% of the total number, compared to 33% of all storms; and 2) during-ESD sampled storms and all storms both comprised 13% of the total number. For storms with 24-hour rainfall exceeding 3.81cm, there was excellent agreement between sampled and all storm frequencies for the pre-ESD period (20% and 19%, respectively), but not for the during ESD period (48% and 20%, respectively; Figure 20).

As an illustration of our sampling strategy and results, we have displayed water quality results for a representative event (event AE: 3/3 – 4/2023, Figures 21 and 22). Total areal storm rainfall was estimated as 1.80cm at PLBR and 2.11cm at UTLP from gage-adjusted NEXRAD data. The PLBR hydrograph displayed the characteristic double-peaked shape, while the UTLP hydrograph was single-peaked. The water quality for both stations revealed: 1) strong dilution of SC, chloride, and sulfate at high discharge; 2) increasing concentrations of TSS, TP, orthophosphate-P, particulate-P, DON, and particulate-N at high discharge; and 3) gradual trending toward antecedent water quality during the

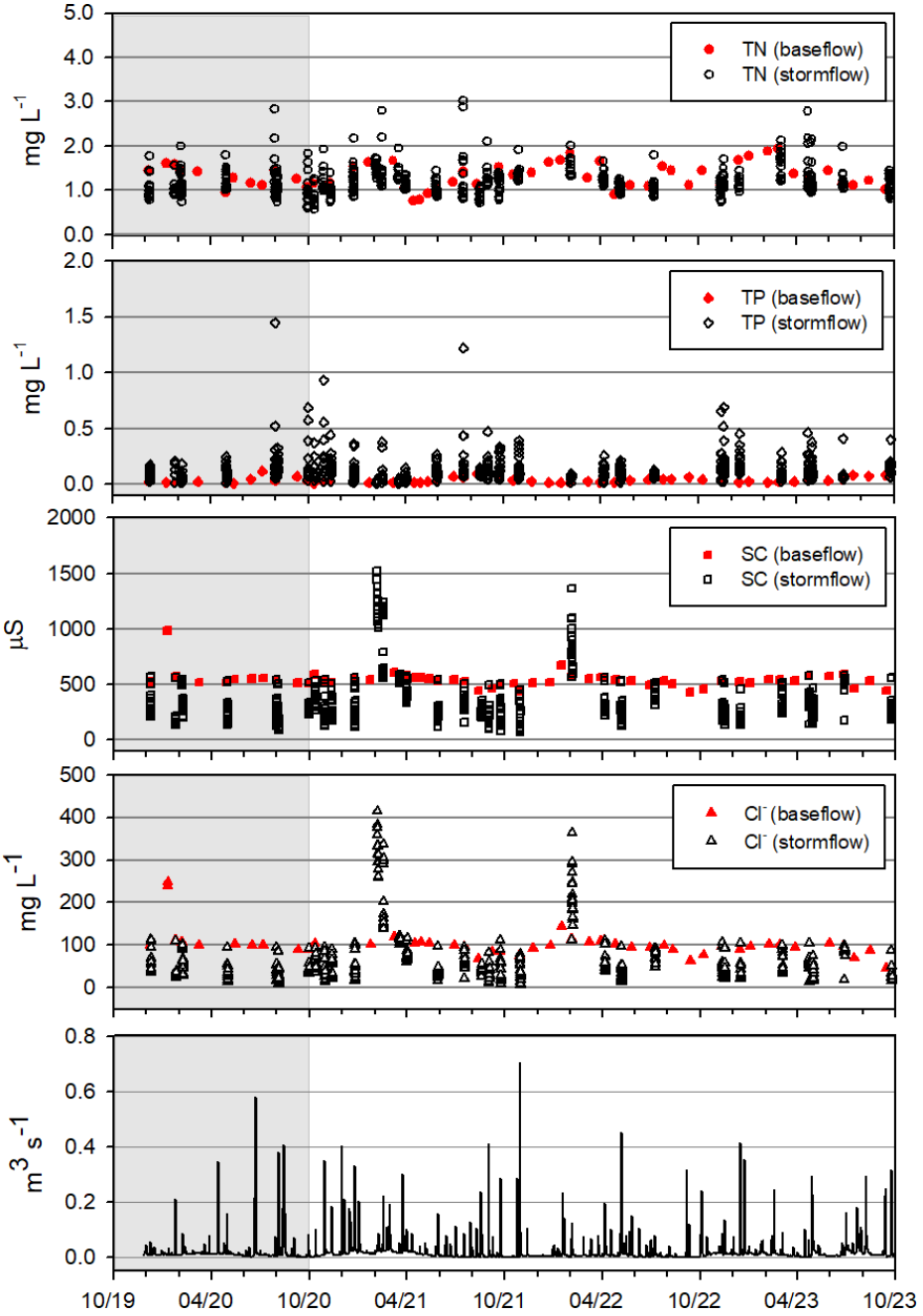

Figure 23. Concentrations of key water quality constituents (TN, TP, SC, and chloride) and discharge during the project at the PLBR station; the pre-ESD period is highlighted in gray.

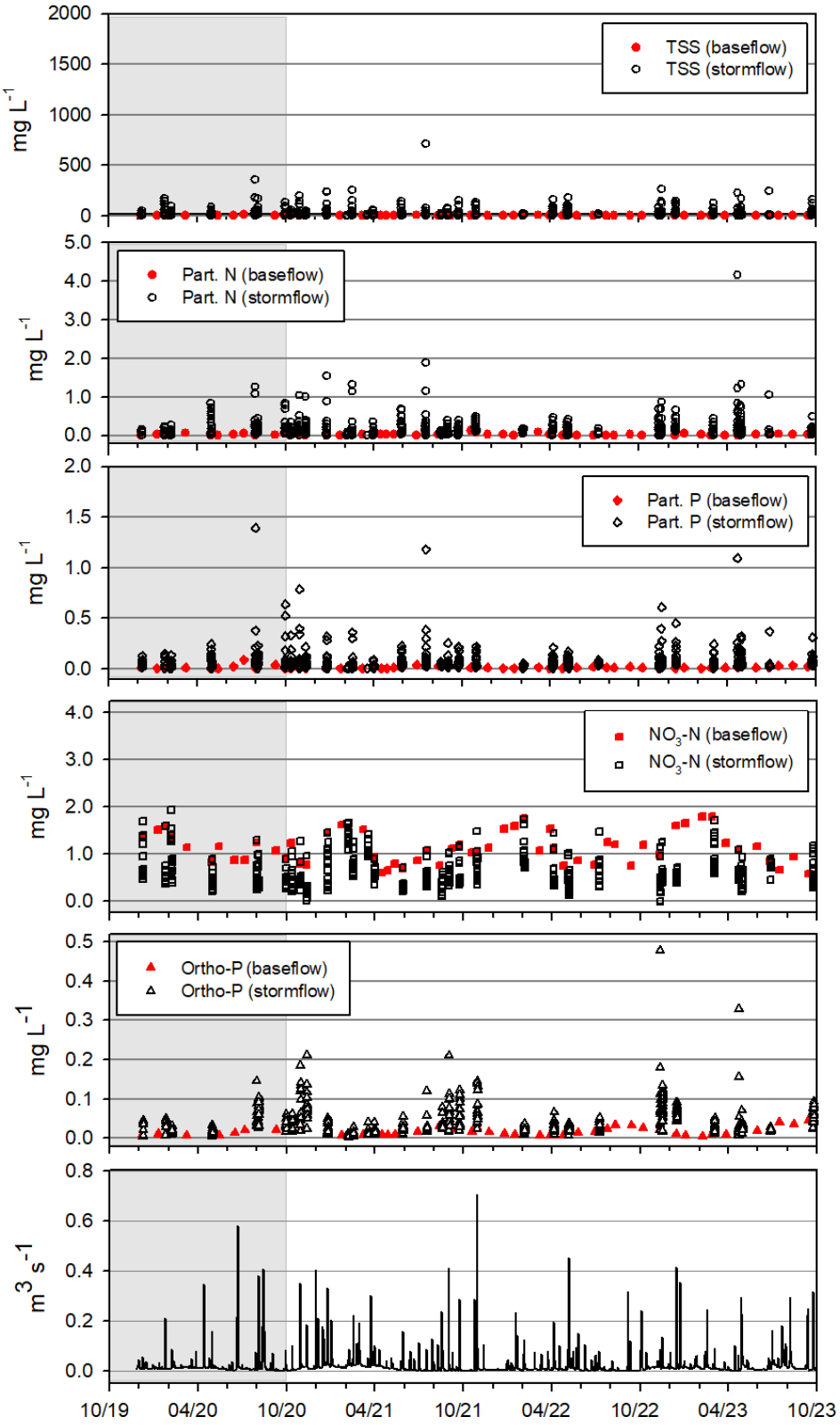

Figure 24. Concentrations of other key water quality constituents (TSS, particulate-N, particulate-P, nitrate-N, and orthophosphate-P) and discharge during the project at the PLBR station; the pre-ESD period is highlighted in gray.

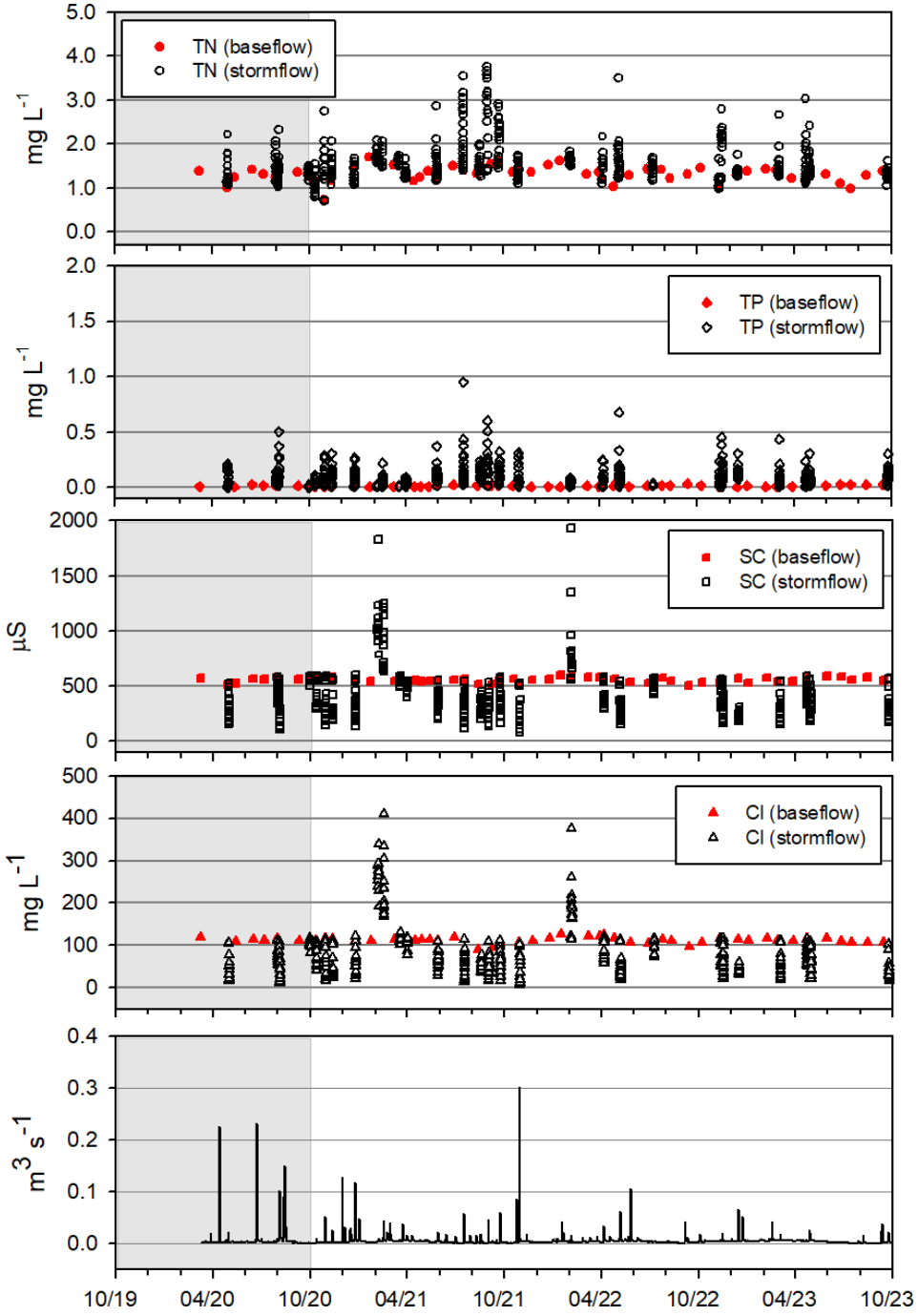

Figure 25. Concentrations of key water quality constituents (TN, TP, SC, and chloride) and discharge during the project at the UTLP station; the pre-ESD period is highlighted in gray.

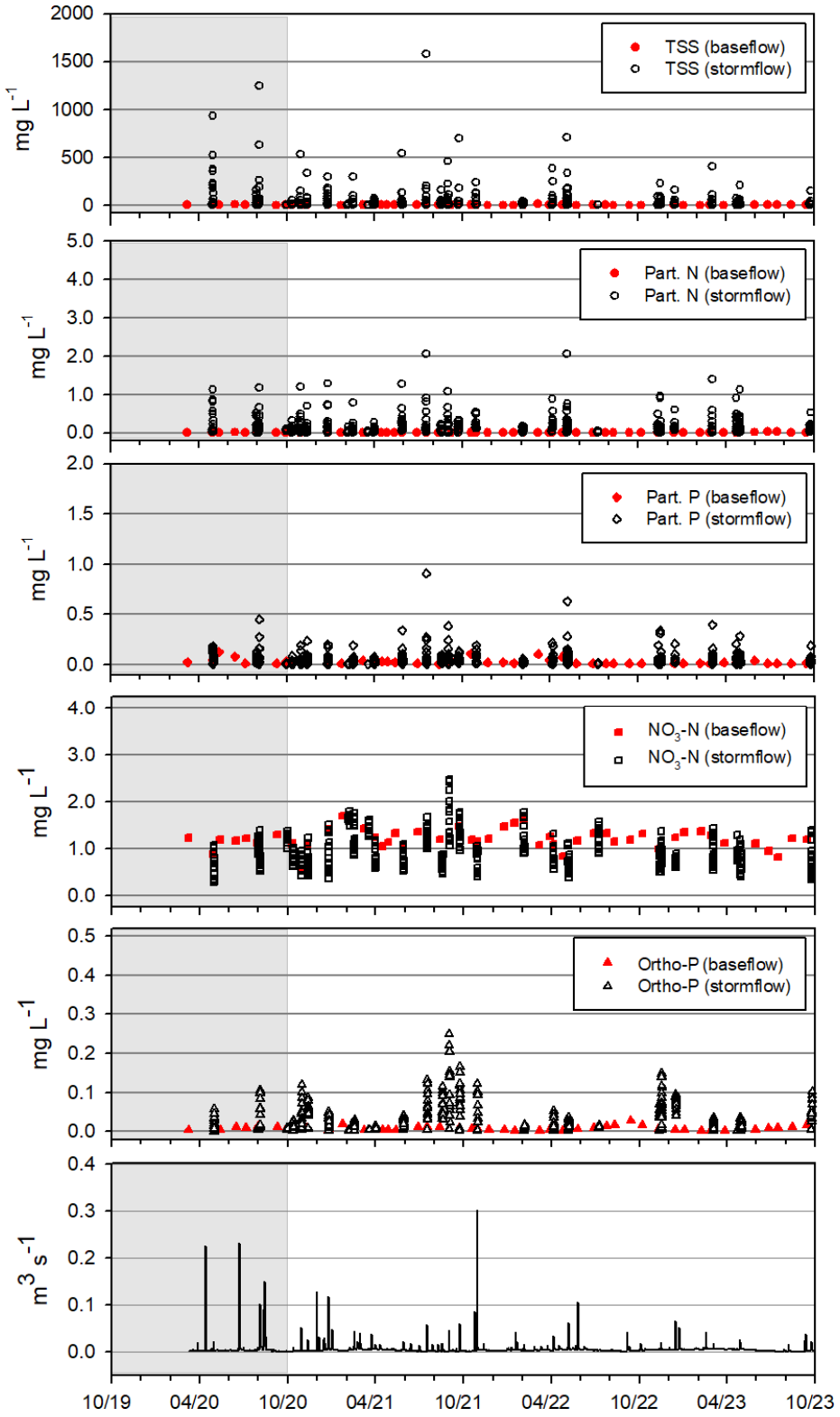

Figure 26. Concentrations of other key water quality constituents (TSS, particulate-N, particulate-P, nitrate-N, and orthophosphate-P) and discharge during the project at the UTLP station; the pre-ESD period is highlighted in gray.

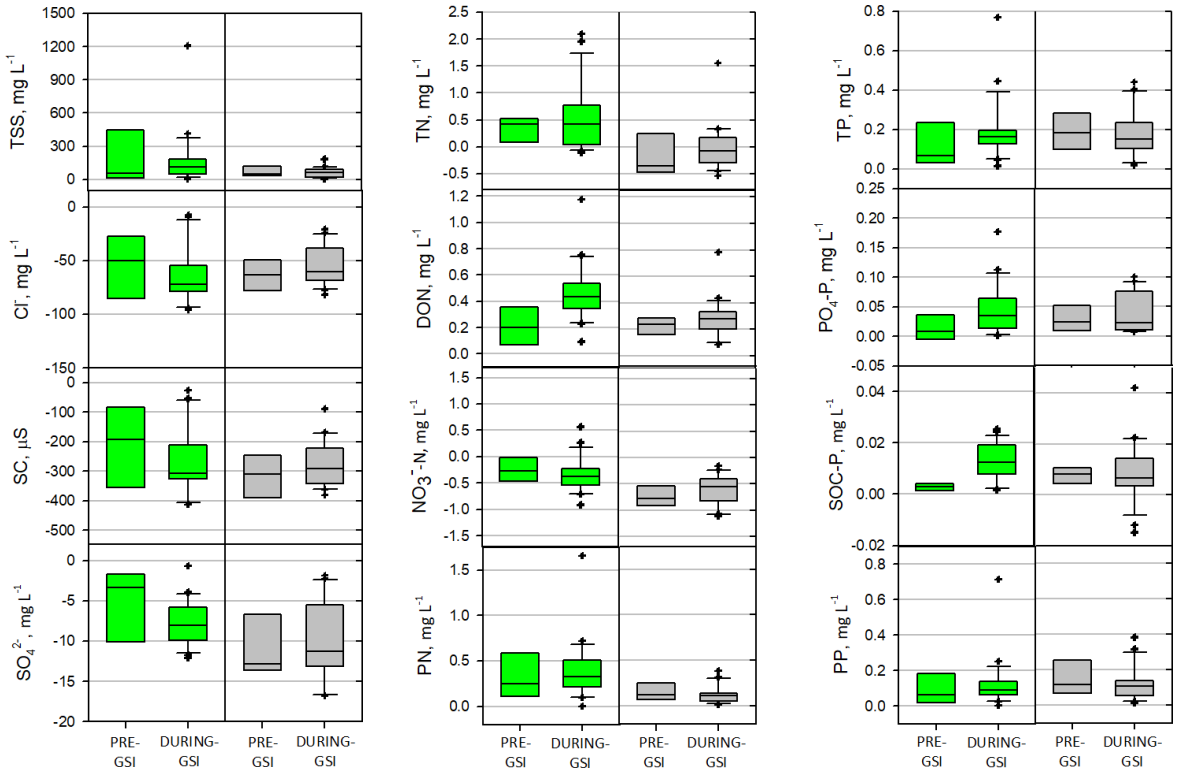

Figure 27. Changes in concentrations (EMC's) of key water quality constituents during stormflow events relative to antecedent baseflow values for the two time periods (pre-ESD and during-ESD); UTLP box-and-whisker plots in green; PLBR box-and-whisker plots in gray.

hydrograph recession. The main difference in water quality response was the behavior of TN and nitrate-N. At PLBR, TN and nitrate-N concentrations both decreased on the rising limb of the event, but stabilized or increased slightly on the falling limb; particulate-N and DON comprised a larger portion of TN near the hydrograph peak, but neither constituent concentration exceeded the nitrate-N concentration (Figure 21). At UTLP, both TN and nitrate-N concentrations increased early in the event, but nitrate declined as the rising limb discharge steepened, while TN continued to increase (as particulate-N, and to a lesser extent, DON, became the dominant forms of N near the peak). On the UTLP falling limb, TN and particulate-N concentrations receded as nitrate-N concentrations gradually rebounded to values similar to those measured at the beginning of the event (~ 1.0 mg N/L; Figure 22).

Importantly, similar patterns as those observed for event AE were very robust across the entire suite of events sampled, although one major difference can be readily discerned. As described previously, SC and chloride concentrations did not always dilute during winter events owing to wash-off of road

deicing chemicals (i.e., salts) that had been applied to paved surfaces within the watersheds prior to these events. This was clearly the case for events "N" (2/15 - 16/2021) and "X" (2/2 – 4/2022) where SC exceeded 1,000 μS and chloride concentrations exceeded 300 mg/L in some samples at both sites. All other responses were at least qualitatively similar to the patterns observed in event AE (Figures 21 and 22). By color-coding baseflow samples (red symbols) and stormflow samples (black symbols), the changes in water quality associated with the stormflow events over the entire project are visually much clearer (Figures 23 - 26).

Another feature evident in the dataset was the relative stability of baseflow water quality during the project. For most of the primary constituents, the temporal variability in concentration was driven almost entirely by stormflow dynamics. This was particularly the case for suspended constituents (e.g., TSS, particulate-N, particulate-P), but also for many of the dissolved (and mixed) constituents (e.g., SC, chloride, TN, TP, and orthophosphate-P). The one exception to this rule was the dynamics of nitrate-N in both streams with each showing considerable interannual variability in baseflow concentrations; especially at PLBR, baseflow nitrate-N concentrations exhibited obvious seasonal variability with highest concentrations (~1.5 mg N/L) in the winter and lowest concentrations (~ 0.5 – 0.8 mg N/L) during the summer growing season (Figures 23 – 26).

The variability in water quality associated with the stormflow events can also be examined by computing the change in concentration of a constituent (i.e., an EMC) as a difference using the antecedent baseflow concentration as a reference. Since baseflow concentrations for most key measured constituents were either very low (e.g., TSS, TP, particulate-N, particulate-P, orthophosphate-P) or stable (e.g., SC and chloride), this approach has the advantage of allowing for both the direction and magnitude of change to be evaluated in a single measurement. We displayed these results as box-and-whisker plots (despite the fact that too few measurements were available for the pre-ESD period to identify any outlier observations). In general, based solely on median values, the results for PLBR displayed much greater stability over time (i.e. between the two periods); exceptions to this rule were the behavior of TN and nitrate-N, in particular, for which differences among the two time periods at UTLP were negligible. Interestingly, these two constituents showed the most inconsistent behavior among events (i.e., some events showed negative chemodynamic behavior, while others showed positive behavior). At UTLP, if a particular constituent exhibited

negative chemodynamic behavior, the median change in that constituent typically became more negative in the during-ESD period; if the constituent showed positive chemodynamic behavior, the median change was more positive in the latter period. Total ranges of the observations were also greatest for UTLP with a couple of exceptions: SOC-P, particulate-P, and sulfate (Figure 27).

Pollutant loads. We used a scientifically well-accepted loading model (LOADEST: Cohn et al. 1989, Runkel et al. 2004) to estimate monthly and annual loads of the various water quality constituents. Virtually all of the seven-parameter load models had very high (80-99%) R^2 values, but quite a few were flagged due to high (> +/- 30%) bias and low (E < 0) model efficiencies (Nash and Sutcliffe 1970); loads from those models are considered unreliable. Loads of several important constituents could not be satisfactorily modeled for either one or both watersheds. Constituents without satisfactory load

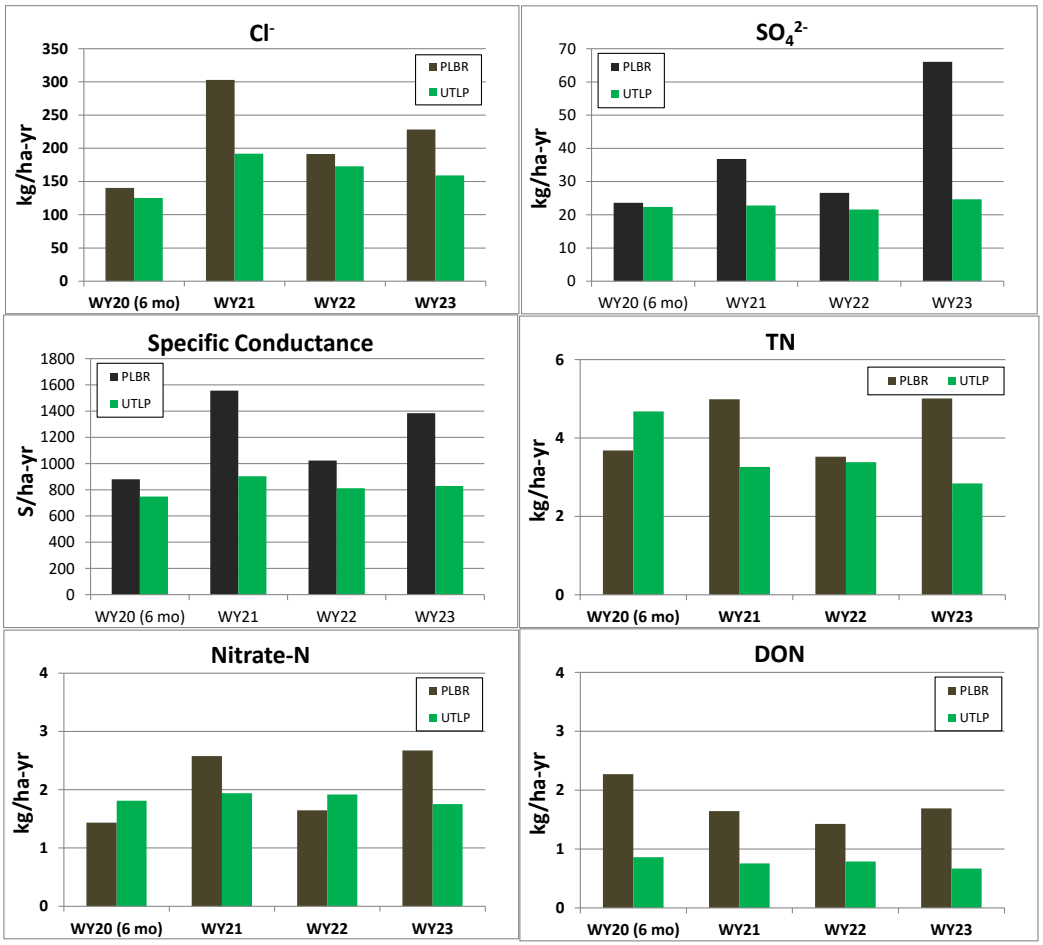

Figure 28. Estimated annual loads of six water quality constituents for the two watersheds based on application of LOADEST; see Tables A3 and A4 for statistical details on the LOADEST models. Since water year 2020 was a half year, the displayed loads for that year were obtained by doubling the estimated loads for the six-month period.

models for either watershed include the following: particulate-N and TSS. In addition, for UTLP, we were unable to obtain satisfactory load models for TP, TDP, particulate-P, orthophosphate-P, and ammonium-N. Undoubtedly, one of the reasons for the difficulties encountered in modeling UTLP loads was the watershed change (including ESD) that occurred during the study. This precluded us from comparing loads for several pollutants—most importantly all of the P forms (Tables A3, A4). We thus focus here only on six key constituents for which satisfactory load models for *both* watersheds were obtained (i.e., chloride, sulfate, SC, TN, nitrate-N, and DON). Annual results for water year 2020 were approximated by doubling the LOADEST loads computed for the period April – September 2020. Annual loads and discharge-weighted concentrations for these constituents are presented in Figures 28 and 29, respectively.

Estimated annual loads were generally most similar between the watersheds in water year 2020 (i.e., the pre-ESD period); this is most apparent for chloride and SC that are least affected by biological activity. The differences in loads of these constituents between the watersheds typically increased over time as well. For chloride, the average annual chloride load for water years 2021 – 2023 at UTLP (175 kg/ha-yr) was about 27% less than at PLBR (241 kg/ha-yr); for SC, the difference was even greater (~36%). With only one exception (nitrate-N in water year 2022), annual loads of all six constituents were higher at PLBR than at UTLP in water years 2021 through 2023. Estimated sulfate, nitrate-N, and DON loads were remarkably stable over time at UTLP. Note that we consider the increases in chloride and SC loads and concentrations in both watersheds between water years 2020 and 2021 to be primarily attributable to the fact that computed loads from water year 2020 lacked the winter road-salting period (Figure 28); the same might also be true of the nitrate-N results given that fact that baseflow nitrate-N concentrations were previously shown to be higher in the winter (Figures 24, 26, and 28). The TN load results may provide the best evidence to support our hypothesis about ESD implementation and water quality. Water year 2020 TN loads were much higher at UTLP than at PLBR despite the fact that runoff from the former watershed was about 17% less (Figure 10); in water years 2021 – 2023, however, the annual average TN load for UTLP (~3.2 kg N/ha-yr) was about 30% lower than at PLBR (4.5 kg N/ha-yr; Figure 28)—tracking or perhaps slightly exceeding the percentage decline in total runoff from the pre-ESD time period (~25%). It is certainly conceivable that ESD implementation reduced particulate-N loads (a portion of the TN loads) at UTLP,

despite the fact that a satisfactory LOADEST model could not be obtained for either watershed (Tables A3, A4).

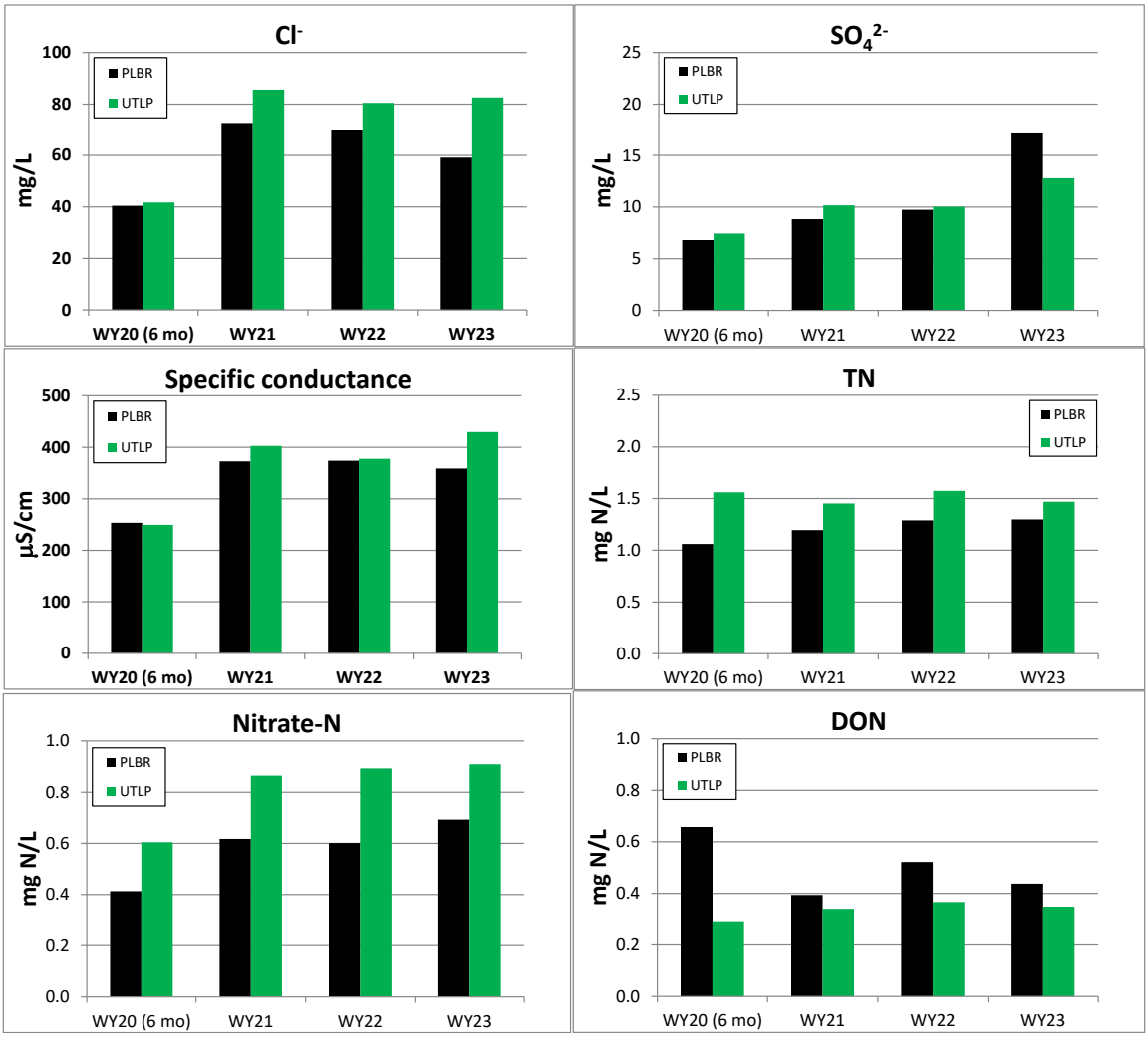

Figure 29. Estimated annual discharge-weighted concentrations of six water quality constituents for the two watersheds based on application of LOADEST; see Tables A3 and A4 for statistical details on the LOADEST models. Note that water year 2020 was a half year, so the displayed values represent the concentrations for the spring and summer of 2020 only.

In terms of annual concentrations, the two least biologically-reactive constituents (chloride, SC) were virtually identical in partial year 2020 (although both values were likely underestimated due to lack of data for winter 2020), but a widening gap between the watersheds during the 2021 – 2023 years is apparent (UTLP > PLBR). Since concentrations of both constituents are higher in baseflow than stormflow (Figure 27), this effect would be consistent with the increasing BFI at UTLP related to ESD implementation (Figure 15). Sulfate concentrations showed a consistent increase in both

watersheds, while nitrate-N concentrations maintained a consistent pattern of higher concentrations at UTLP over the project; DON concentrations displayed the opposite pattern (PLBR > UTLP). Again, the TN results conceivably provide the best evidence for an ESD effect: the UTLP TN concentration in water year 2020 (1.6 mg N/L) was about 40% higher than at PLBR (1.1 mg N/L) but in water years 2021 through 2023 the difference was less than 10% (Figure 29).

Discussion

Despite the fact that the project allowed for a very short (~6-months) pre-treatment period, the monitoring data provided very strong support for our hypothesis that ESD implementation in the UTLP watershed progressively reduced total runoff, storm event runoff, and storm event runoff ratios as others have documented. Total runoff was shown to decrease by about 42% relative to PLBR, but this is likely an overestimate of the overall effect given the estimated 17% difference based on data from the pre-ESD period. The difference for the pre-ESD period could be explained by one or more factors including uncertainties in delineating the watersheds and a real difference in evapo-transpirative demand due to greater forest cover in UTLP. Therefore, our best estimate for the ESD effect on total runoff was a decrease of about 25%. This reduction in total runoff is actually about a factor of two larger than the percentage of the UTLP watershed that was disturbed (~13%) during the most recent development sequence—suggesting that ESD has had an outsized impact on total runoff. There are several possible mechanisms that could explain the disproportionate effect of ESD BMP's.

The first mechanism is enhanced groundwater recharge (commonly called "exfiltration" by engineers) associated with both the bioretentions and the dry wells. Since these features are un-lined, the (positive) head gradient between these facilities and the surrounding soil media would likely be enhanced as they fill during storms; the drier surrounding media could be expected to exert a sorptive effect on water detained in the bioretentions and dry wells—thus producing actual stormwater retention at the watershed scale. Of course, if the surrounding media drains to gaged watershed outlet points, the enhanced recharge would likely be "captured" as increased baseflow. In this case, the ESD facilities would be playing a stormwater detention role only. If, on the other hand, the enhanced recharge contributes to what is sometimes called "deep seepage" by hydrologists, then the ESD effect on runoff would likely be greater (i.e., the recharge would be considered completely

retained). This latter effect appears to be what has occurred at UTLP in response to ESD implementation.

A second possibility is that the enhanced infiltration into the bioretentions has fueled an increase in evapotranspirative losses of water; this could occur in the bioretention facilities themselves or, more likely, from areas to which the bioretentions are discharging water. In UTLP, these outlets were set back considerable distances (~20m or more) from the stream itself, thus providing an opportunity for discharged water to re-infiltrate soils within the floodplain and support evapotranspiration by floodplain vegetation. Hydrologists have coined the term "runoff-run-on" behavior to describe this mechanism. It should also be emphasized that the recent UTLP development largely avoided disturbances of the stream channel itself and provided strong protection (i.e., a buffer) of the forested floodplain. Anecdotally, we observed minimal disturbance of the forest vegetation along the UTLP mainstem or tributaries.

A final possibility is that the new UTLP development altered watershed boundaries or resulted in the diversion of stormwater off of the watershed through an alternate pathway. In support of the diversion hypothesis, site development plans that we obtained for the recent development within the UTLP watershed (and in adjacent watersheds) from Howard County revealed that a small portion of UTLP stormwater could possibly be diverted to a regional stormwater pond. We do not believe that the observed changes in total runoff could be fully explained by the latter mechanisms, but we cannot completely rule them out as contributing factors.

The statistically-significant changes in storm event runoff and runoff ratios (quantified as median differences between PLBR and UTLP) associated with ESD implementation are at least qualitatively consistent with the observed decrease in total runoff. These changes also appear large given the relatively small portion of the watershed that was developed during the monitoring project, but the way in which these changes were calculated (i.e., as differences between watersheds) makes it difficult to relate them directly to land surface change metrics. Moreover, all of these effects are consistent with another robust effect that we managed to capture using the optimized RDF approach underpinned by an extensive set of chemical hydrograph separations: a monotonic increasing trend in baseflow index (BFI) in UTLP with no trend noted for the control watershed. It is important to note here that we have not observed dramatic increases in baseflow volumes in the UTLP watershed. In

fact, annual baseflow runoff does not appear to have changed much during the study: baseflow volumes were particularly stable in full water years 2021 through 2023. Conversely, direct runoff volumes declined by about 50% from 2021 (110mm) to 2021 (55mm)—dominating the observed BFI signal. At least a portion of the observed decline in direct runoff (and increase in BFI) could have been due to lower precipitation in water years 2022 and 2023 relative to 2021 (as we observed at PLBR), but we consider it unlikely that hydroclimatic variability alone could explain this phenomenon. As with the other metrics, the major caveat in explaining the observed hydrologic changes is the brevity of the pre-treatment period and the fact that it was less than a complete water year. A second issue is the lack of data for a "post-ESD" period since monitoring for this phase of the project ended in September 2023 prior to the completion of ESD implementation.

Our ability to detect changes in peak event runoff and maximum new water contributions provides additional support for our hypotheses and mechanistic understanding. The observed changes in each of these metrics is consistent with the hypothesis that ESD can attenuate stormwater runoff by reducing overland flow contributions. If ESD implementation effectively reduced storm event runoff and storm event ratios (i.e., volumetric metrics), it should not be surprising to observe comparable effects on stormwater intensity. The real question here is whether the observed intensity effects are simply another manifestation of the volumetric effects, since reducing storm event runoff volumes while maintaining the hydrograph shape would be expected to proportionally reduce peak runoff and overland flow (new water) contributions. One way of addressing this question is through application of the unit hydrograph approach, but as we discussed earlier, our analyses to-date that were based on two or three storms are wholly inadequate for this purpose. We suggest here that unit hydrograph analysis might be performed over a longer period of watershed change (i.e., from pre-ESD through post-ESD implementation), but current approaches that attempt to generate a single composite unitgraph from multiple storms would clearly be unsuitable for this type of analysis.

While the hydrologic metrics provided strong and robust support for changes attributable to ESD implementation at the watershed scale, the water quality results were much less conclusive. Despite analyzing nearly 1200 water samples (18 constituents) and utilizing several different approaches to address the research hypotheses, we were unable to detect any significant changes in water quality during the 3.5-year project. Baseflow water quality proved particularly stable in both watersheds—

presumably due to long subsurface residence times that effectively slows the pace of geochemical evolution. Stormflow was obviously quite dynamic, but extreme hydroclimatic variability apparently overwhelmed the impacts that ESD implementation might have had on water quality and loads. One problem that we identified earlier was the extremely short pre-ESD period. This problem turned out to be even more consequential for the water quality part of the study than for the hydrologic component. We simply were unable to characterize water quality variations for a sufficient number of storms during the pre-ESD period of the project to provide a tighter baseline characterization for the latter part of the study. A second issue is that despite strong support in the literature, the loading model (LOADEST) that we chose to work with may not be (in hindsight) the best choice for watersheds that are exhibiting non-stationary hydrologic behavior. The most important independent parameter in the LOADEST models was discharge at the time of sample collection. Obviously if discharge is exhibiting non-stationary behavior, then it is unlikely that the relationship between discharge and load would be stationary. The role of non-stationarity would also be consistent with the fact that we were able to obtain satisfactory loads for 16 of the 18 constituents measured for the PLBR watershed, but only ten of the 18 constituents for UTLP (Tables A3, A4). For future load modeling work, we propose using the newer WRTDS load model that may be inherently better suited for addressing water quality variations in watersheds disturbed by urbanization or other rapid land use changes.

While we obviously cannot go back and increase the length of the monitored pre-treatment period, we are committed to continuing the monitoring effort in both watersheds for several additional years (i.e., collecting equivalent data through a “post-ESD” period). Assuming no other changes in land use in the study watersheds occur in the immediate future, our expectation is that data for the post-ESD period would strongly reinforce the hydrologic and water quality changes that emerged during the current project. Future monitoring would build on the data foundation provided by the current project, as well as allow us to test some other data analysis methods. Fortunately, we were able to secure additional funding from Chesapeake Bay Trust Restoration Research Program that will allow for the overall project to continue through water year 2025. We will be relying on the monitoring systems that remain in place in these Howard County watersheds to extend the project into the post-development phase.

Acknowledgments

The principal investigator gratefully acknowledges the contributions of the following organizations and individuals to the overall success of this project:

- Chesapeake Bay Trust Restoration Research Program, Maryland DNR, and Maryland DOT State Highway Administration: research sponsorship and support from Sadie Drescher (CBT project officer) and Ari Engelberg (DNR)
- Howard County (Mark Richmond *et al.*): right-of-entry permits, site development plans
- Briana Rice (AL): field and laboratory assistance
- Katie Kline (AL): laboratory assistance, quality assurance, and data management
- Jim Garlitz (AL): laboratory assistance
- Joel Bostic (AL): GIS and mapping assistance; load calculations
- Ev Demott (AL): lab assistance
- Trevor Frissell (AL): lab assistance
- Elizabeth Eshleman: field assistance
- Neal Eshleman: field assistance

References Cited

APHA. 2017. *Standard Methods for the Examination of Water and Wastewater*, 23rd Edition. American Public Health Association, Washington, DC.

Barrett, M.E. 2008. Comparison of BMP Performance Using the International BMP Database. *J. Irrig. Drain. Eng.* ***134***:556-561.

Bedan, E.S., and J.C. Clausen. 2009. Stormwater runoff quality and quantity from traditional and low impact development watersheds. *J. Amer. Water Resour. Assoc.* ***45***:998-1008.

Biohabitats, Inc. 2019. Plumtree Branch watershed stormwater retrofit study; https://www.howardcountymd.gov/environmental-services/resource/plumtree-branch-stormwater-retrofit-study-final-report-june-2019.

Bonnin, G.M., D. Martin, B. Lin, T. Parzybok, M. Yekta, and D. Riley. 2004. NOAA Atlas 14, Precipitation-Frequency Atlas of the United States. NOAA, Silver Spring, MD.

Cohn, T.A., L.L. Delong, E.J. Gilroy, R.M. Hirsch, and D.K. Wells. 1989. Estimating constituent loads. *Water Resour. Res. 25*:937-942.

Davis, A.P., W.F. Hunt, R.G. Traver, and M. Clar. 2009. Bioretention technology: Overview of current practice and future needs. *J. Environ. Eng.* ***135***:109-117.

Dietz, M.E. 2007. Low impact development practices: a review of current research and recommendations for future directions. *Water, Air Soil. Pollut.* ***186***:351-363.

Eckhardt, K. 2005. How to construct recursive digital filters for baseflow separation. *Hydrol. Proc.* ***19***:507-515.

Eckhardt, K. 2008. A comparison of baseflow indices, which were calculated with seven different baseflow separation methods. *J. Hydrol.* ***352***:168-173.

Eshleman, K.N., J.S. Pollard, and A.K. O'Brien. 1993. Determination of contributing areas for saturation overland flow from chemical hydrograph separations. *Water Resour. Res.* ***29***:3577-3587.

Fishman, M.J. 1993. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—determination of inorganic and organic constituents in water and fluvial sediments. *U.S. Geological Survey Open-File Report 93-125.*

Foks, S.S., J.P. Raffensperger, C.A. Penn, and J.M. Driscoll. 2019. Estimation of base flow by optimal hydrograph separation for the conterminous United States and implications for national-extent hydrologic models. *Water* ***11***:1629; doi:10.3390/w11081629.

Hewlett, J.D., H.W. Lull, and K.G. Reinhart. 1969. In defense of experimental watersheds. *Water Resour. Res*. ***5***:306-316.

Hornbeck, J.W., R.S. Pierce, and C.A. Federer. 1970. Streamflow changes after forest clearing in New England. *Water Resources Research* ***6***:1124-1132.

Hood, M.J., J.C. Clausen, and G.S. Warner. 2007. Comparison of stormwater lag times for low impact and traditional residential development. *J. Amer. Water Resour. Assoc.* ***43***:1036-1046.

Hopkins, K.G., N.B. Morse, D.J. Bain, N.D. Bettez, N.B. Grimm, J.L. Morse, ad M.M. Palta. 2015. Type and timing of stream flow changes in urbanizing watersheds in the eastern U.S. *Elementa: Sci. Anthropocene* ***3***:000056; doi:10.12953/journal.elementa.000056.

Hribar, A., and C. Lyons. 2017. 2017 Valley Mede Drainage Study: Plumtree Branch and Little Plumtree Branch. Report prepared by McCormick Taylor for Maryland Dept. of Transportation State Highway Administration and Howard County Stormwater Management Division.

Jarden, K.M., A.J. Jefferson, and J.M. Grieser. 2016. Assessing the effects of catchment-scale urban green infrastructure retrofits on hydrograph characteristics. *Hydrol. Proc.* ***30***:1536-1550.

Liu, J., D.J. Sample, C. Bell, and Y. Guan. 2014. Review and research needs of bioretention used for treatment of urban stormwater. *Water* ***6***:1069-1099.

Loperfido, J.V., G.B. Noe, S.T. Jarnagin, and D.M. Hogan. 2014. Effects of distributed and centralized stormwater best management practices and land cover on urban stream hydrology at the catchment scale. *J. Hydrol.* ***519***:2584-2595.

Lott, D.A., and M.T. Stewart. 2016. Base flow separation: A comparison of analytical and mass balance methods. *J. Hydrol.* ***535***:525-533.

Kratky, H., Z. Li, Y. Chen, C. Wang, X. Li, and T. Yu. 2017. A critical literature review of bioretention research for stormwater management in cold climate and future research recommendations. *Front. Environ. Sci. Eng.* ***11***:1-15.

Nash, J.E., and J.V. Sutcliffe. 1970. River flow forecasting through conceptual models part I — A discussion of principles. *J. Hydrol.* ***10*** (3):282–290.

O'Driscoll, M., S. Clinton, A. Jefferson, A. Manda, and S. McMillan. 2010. Urbanization effects on watershed hydrology and in-stream processes in the southern United States. *Water **2***:605-648.

Pellerin, B.A., W.M. Wollheim, X. Feng, and C.J. Vorosmarty. 2008. The application of electrical conductivity as a tracer for hydrograph separation in urban catchments. *Hydrol. Proc. **22***:1810-1818.

Rose, S., and N.E. Peters. 2001. Effects of urbanization on streamflow in the Atlanta area (Georgia, USA): a comparative hydrological approach. *Hydrol. Proc. **15***:1441-1457.

Runkel, R.L., C.G. Crawford, and T.A. Cohn. 2004. Load Estimator (LOADEST): a FORTRAN Program for Estimating Constituent Loads in Streams and Rivers. *Techniques and Methods Book 4*, Chapter A5. U.S. Geological Survey, Reston, VA.

Scarlett, R.D., S.K. McMillan, C.D. Bell, S.M. Clinton, A.J. Jefferson, and P.S.C. Rao. 2018. Influence of stormwater control measures on water quality at nested sites in a small suburban watershed. *Urban Water J. **15***:868-879.

Schuster, W., and L. Rhea. 2013. Catchment-scale hydrologic implications of parcel-level stormwater management (Ohio USA). *J. Hydrol. **485***:177-187.

Selbig, W.R., and R.T. Bannerman. 2008. A comparison of runoff quantity and quality from two small basins undergoing implementation of conventional- and low-impact-development (LID) strategies: Cross Plains, Wisconsin, water years 1999-2005. *U.S. Geological Survey Scientific Investigations Report 2008-5008.*

Smith, J.A., M.L. Baeck, G. Villarini, C. Welty, A.J. Miller, and W.F. Krajewski. 2012. Analyses of long-term, high-resolution radar rainfall data set for Baltimore metropolitan region. *Water Resour. Res. **48***:W04504.

USEPA. 1993. Methods for the Determination of Inorganic Substances in Environmental Samples. *EPA/600/R-93/100*, U.S.EPA National Exposure Research Laboratory (NERL) Microbiological and Chemical Exposure Assessment Research Division (MCEARD), Cincinnati, OH.

Walsh, C.J., A.H. Roy, J.W. Feminella, P.D. Cottingham, P.M. Groffman, and R.P. Morgan. 2005. The urban stream syndrome: current knowledge and a search for the cure. *J. North Amer. Benth. Soc. **24***: 706-723.

Zhang, R., Q. Li, T.L. Chow, S. Li, and S. Danielescu. 2013. Baseflow separation in a small watershed in New Brunswick, Canada, using a recursive digital filter calibrated with the conductivity mass balance method. *Hydrol. Proc. **27***:2659–2665.

Appendices

Table A1. Current stormwater management facilities in the UTLP watershed as of October 1, 2023.

Plan No. or BMP Name	SWM Code (no.)	SWM Type	Drainage Area (ft^2)	ISA (ft^2)	Notes
N/A	N/A (10)	2A Grass Swales (abandoned)	663,600	434,958	I-70 legacy ISA (areas est.); SWM from MDOT-SHA NPDES SWMFAC
F-88-232_POND	N/A (1)	Wet Pond	1,533,748	744,466	MDE StormwaterPrint; ISA est.
F-87-188_POND	N/A (1)	Dry Pond	217,800	105,718	MDE StormwaterPrint; ISA est.
F-93-073_POND	N/A (1)	Wet Pond	530,125	257,317	MDE StormwaterPrint; ISA est.
F-07-158	F-6 (2)	Bioretention	28,750	28,750	Resort Road extension #1 (areas est.)
F-16-004	F-6 (1)	Bioretention	43,560	43,560	Resort Road extension #2 (areas est.)
F-17-095	M-6 (2)	Microbioretention	40,904	22,156	Areas from development plan
F-17-095	F-6 (1)	Bioretention	60,657	32,753	Areas from development plan
F-17-096	F-6 (5)	Bioretention	249,205	136,610	Areas from development plan
F-17-096	M-6 (1)	Microbioretention	16,200	5,139	Areas from development plan
F-17-096	M-5 (44)	Dry wells	35,640	35,640	Areas est. from development plan (44 x 810)
F-18-027	M-6 (1)	Microbioretention	22,684	12,906	Areas from development plan
F-18-027	M-5 (14)	Dry wells	11,550	11,550	Areas from development plan
SDP-20-036	M-6 (7)	Microbioretention	131,983	123,042	Areas from development plan
Totals (acres), 2015			67.6	35.4	
Totals (acres), 2023			82.3	45.8	
Watershed area (acres)			198.4		
ISA (%), 2015				**17.8**	
ISA (%), 2023				**23.1**	~30% increase in ISA (2015 – 2023)

Table A2. Major stormflow-producing rainfall events. Recurrence intervals (R.I.) taken from G.M. Bonnin *et al.* (2004; NOAA Atlas 14, Point precipitation frequency estimates: Ellicott City, MD). Water quality samples were collected and analyzed for events with lettered storm ID's. Data for events "M" and "P" excluded.

Storm ID	Date(s)	PLBR rainfall-gage adj. NEXRAD (cm)	UTLP rainfall-gage-adj. NEXRAD (cm)	1-hr gage rainfall (cm)	24-hr gage rainfall (cm)	1-hr R.I. (yr)	24-hr R.I. (yr)
"A"	12/9-11/19	N/A	N/A	0.53	2.01	<1	<1
"B"	1/25-26/20	3.96	3.73	0.89	3.48	<1	<1
"C"	2/6-8/20	2.77	3.28	0.33	2.31	<1	<1
N/A	3/18-19/20	1.52	1.30	0.41	1.70	<1	<1
N/A	03/28/20	2.24	2.13	0.74	1.83	<1	<1
N/A	4/12-14/20	6.22	6.45	1.02	5.46	<1	<1
N/A	4/23-24/20	2.13	2.03	0.36	2.11	<1	<1
N/A	04/26/20	1.22	1.27	0.53	1.35	<1	<1
"D"	4/30 - 5/1/20	N/A	N/A	0.71	3.48	<1	<1
N/A	6/4-5/20	2.67	1.75	0.86	2.84	<1	<1
N/A	6/10-11/20	1.52	1.45	0.71	3.48	<1	<1
N/A	6/20-21/20	3.35	2.31	4.62	4.90	**5**	<1
N/A	6/22-23/20	6.30	5.49	5.97	6.07	**20**	<1
"E"	7/30-31/20	N/A	N/A	2.31	3.12	<1	<1
"F"	8/3-4/20	N/A	N/A	1.55	8.00	<1	**2**
N/A	8/12-13/20	N/A	N/A	1.14	2.84	<1	<1
N/A	8/14-15/20	6.27	8.00	2.36	2.82	<1	<1
N/A	8/28-29/20	2.06	2.13	0.46	1.80	<1	<1
N/A	9/03-04/20	0.74	0.76	1.22	1.35	<1	<1
"G"	9/29-30/20	N/A	N/A	1.65	2.39	<1	<1
"H"	10/11-12/20	3.00	2.84	0.61	2.54	<1	<1
"J"	10/29-30/20	5.23	4.37	1.19	5.66	<1	<1
"K"	11/11-12/20	5.23	5.11	0.94	4.88	<1	<1
N/A	11/30 - 12/1/20	4.34	3.73	1.73	4.57	<1	<1
"L"	12/24-25/20	4.24	3.81	0.89	4.57	<1	<1
N/A	1/1-2/21	3.07	2.90	0.84	3.28	<1	<1
"N"	2/15-16/21	2.69	2.57	0.41	2.54	<1	<1
N/A	3/24-25/21	3.91	3.07	1.63	4.78	<1	<1
"Q"	3/31 - 4/1/21	2.26	2.39	0.66	2.39	<1	<1
"R"	5/28-29/21	4.65	4.27	2.26	4.29	<1	<1
"S"	7/17-18/21	1.50	2.82	0.64	1.07	<1	<1
"T"	8/17-19/21	5.54	5.26	0.79	4.37	<1	<1
N/A	08/20/21	1.30	1.14	0.46	1.40	<1	<1
"U"	09/01/21	4.24	3.68	0.71	4.75	<1	<1
"V"	09/23/21	2.87	2.92	1.80	3.05	<1	<1
N/A	10/25-26/21	6.10	6.60	1.91	5.72	<1	<1
"W"	10/29-30/21	6.83	6.76	1.83	7.14	<1	**1**
N/A	1/16-17/22	2.49	2.49	1.07	2.44	<1	<1
"X"	2/2-4/22	2.82	2.44	0.25	1.96	<1	<1
"Y"	4/5-6/22	2.97	2.79	0.64	3.15	<1	<1
N/A	4/18-19/22	2.46	2.34	0.81	2.69	<1	<1
"Z"	5/6-8/22	6.88	6.25	0.94	5.36	<1	<1
N/A	5/22-23/2022	1.57	1.85	0.71	1.65	<1	<1
N/A	5/27-28/22	2.84	3.07	1.88	2.51	<1	<1
N/A	6/08-09/22	3.43	3.43	2.13	2.79	<1	<1
N/A	6/22-23/22	1.27	1.42	0.20	1.32	<1	<1
"AA"	7/08-10/22	1.27	1.09	0.28	1.40	<1	<1
N/A	7/18-19/22	0.64	0.36	0.41	0.64	<1	<1
N/A	7/31/-8/1/22	0.81	0.86	0.20	0.46	<1	<1
N/A	8/04-05/22	1.12	0.99	0.56	1.04	<1	<1
N/A	8/05-06/22	0.81	1.04	0.61	0.74	<1	<1
N/A	08/15/22	0.76	0.66	0.30	0.69	<1	<1
N/A	8/17-18/22	0.53	0.33	1.32	1.35	<1	<1
N/A	9/5-7/222	6.27	6.12	2.24	6.27	<1	<1
N/A	9/11-12/22	2.06	2.31	0.97	2.36	<1	<1
N/A	9/30-10/1/22	2.18	1.98	0.69	2.11	<1	<1

Table A2 (continued). Major stormflow-producing rainfall events.

Storm ID	Date(s)	PLBR rainfall-gage adj. NEXRAD (cm)	UTLP rainfall-gage-adj. NEXRAD (cm)	1-hr gage rainfall (cm)	24-hr gage rainfall (cm)	1-hr R.I. (yr)	24-hr R.I. (yr)
N/A	10/02-04/22	4.65	4.34	0.66	2.59	<1	<1
N/A	10/04-05/22	0.43	0.33	0.18	0.38	<1	<1
N/A	10/13-14/22	1.73	2.31	0.51	1.70	<1	<1
N/A	10/17-18/22	0.38	0.46	0.30	0.41	<1	<1
"AB"	11/11-12/22	2.36	2.24	0.86	2.34	<1	<1
"AC"	11/15/16/22	2.67	2.69	0.69	2.72	<1	<1
N/A	11/27/22	0.51	0.51	0.36	0.53	<1	<1
N/A	11/30 - 12/01/22	0.79	0.74	0.28	0.84	<1	<1
N/A	12/03-04/22	1.12	1.19	0.56	1.17	<1	<1
"AD"	12/15-16/22	7.04	6.48	0.66	6.91	<1	<1
N/A	12/22-23/22	5.05	4.57	0.76	5.05	<1	<1
N/A	12/31/22-1/01/23	0.86	0.79	0.25	0.86	<1	<1
N/A	1/05-06/23	0.20	0.13	0.20	0.36	<1	<1
N/A	1/12-14-23	0.86	0.69	0.18	0.66	<1	<1
N/A	1/19-20/23	0.56	0.58	0.28	0.51	<1	<1
N/A	1/22-24/23	1.04	0.91	0.15	1.12	<1	<1
N/A	1/25-26/23	1.32	1.32	0.48	1.50	<1	<1
N/A	2/12-14/23	2.46	2.29	0.58	2.39	<1	<1
N/A	2/16-18/23	3.68	3.81	0.89	3.68	<1	<1
N/A	2/27-28/23	0.84	0.84	0.36	0.91	<1	<1
"AE"	3/3-4/23	1.80	2.11	0.48	1.65	<1	<1
N/A	3/10-11/23	0.43	0.30	0.18	0.41	<1	<1
N/A	3/23-26/23	1.83	1.70	0.46	1.30	<1	<1
N/A	4/01-02/23	0.74	0.91	0.18	0.58	<1	<1
N/A	4/15-16/23	1.83	1.85	1.32	1.80	<1	<1
"AF"	4/22-23/23	2.54	1.70	1.30	1.68	<1	<1
"AG"	4/28/23 - 5/1/23	6.81	6.43	1.09	4.78	<1	<1
N/A	5/13-14/23	0.91	0.89	0.36	0.91	<1	<1
N/A	5/20-21/23	0.53	0.53	0.43	0.53	<1	<1
N/A	6/21-24/23	2.24	2.24	0.33	0.97	<1	<1
"AH"	6/27-28/23	0.51	0.18	0.79	0.86	<1	<1
N/A	6/30/23 - 7/1/23	0.92	0.95	0.91	0.97	<1	<1
N/A	7/1-2/2023	1.37	1.03	1.60	1.68	<1	<1
N/A	7/14-15/23	0.97	0.57	1.52	1.55	<1	<1
N/A	7/21-22/23	1.80	1.43	1.93	2.26	<1	<1
N/A	7/24-25/23	0.63	0.54	0.51	0.79	<1	<1
N/A	7/28-29/23	2.34	2.67	1.73	1.78	<1	<1
N/A	8/6-7/23	1.86	1.92	0.74	1.60	<1	<1
N/A	8/7-8/23	2.85	3.17	2.39	2.64	<1	<1
N/A	8/10-11/23	0.88	0.93	0.05	0.05	<1	<1
N/A	9/08-09/23	1.33	0.74	0.66	0.89	<1	<1
N/A	9/10-12/23	2.28	3.71	1.19	2.31	<1	<1
N/A	9/12-13/23	3.04	3.12	2.11	3.20	<1	<1
N/A	9/17-18/23	0.60	0.60	0.23	0.61	<1	<1
"AJ"	9/22-25/23	5.47	5.18	1.07	3.84	<1	<1

Table A3. Statistical output from LOADEST models developed from the project water quality data for PLBR (water years 2020 – 2023). Constituents for which the percent bias, B_P, exceeded ±30% and/or the Nash-Sutcliffe efficiency, E, was less than zero were considered unreliable and were not interpreted (results highlighted in yellow).

Constituent	R^2	PPCC	B_P	PLR	E
Chloride	93.8	0.981	-3.1	0.969	0.459
SC	96.9	0.975	0.83	1.008	0.683
TSS	95.0	0.996	27.1	1.271	-0.108
Sulfate	96.0	0.985	-1.7	0.983	0.544
TP	96.8	0.984	15.4	1.154	0.424
TDP	97.3	0.972	11.8	1.118	0.866
Particulate-P	94.0	0.990	19.1	1.191	0.026
Orthophosphate-P	96.6	0.984	7.88	1.079	0.882
SOC-P	95.8	0.977	14.9	1.149	0.532
TN	98.7	0.961	-5.3	0.947	0.864
TDN	98.8	0.984	-2.4	0.976	0.900
Particulate-N	89.5	0.995	79.5	1.794	-2.89
Nitrate-N (IC)	94.6	0.940	-3.7	0.963	0.701
Nitrate-N (FIA)	95.8	0.988	-4.1	0.959	0.713
Nitrite-N	93.4	0.988	3.16	1.032	0.485
Ammonium-N	84.8	0.993	11.9	1.119	0.254
DON	97.2	0.962	26.4	1.264	0.371

Table A4. Statistical output from LOADEST models developed from the project water quality data for UTLP (water years 2020 – 2023). Constituents for which the percent bias, B_P, exceeded ±30% and/or the Nash-Sutcliffe efficiency, E, was less than zero were considered unreliable and were not interpreted (results highlighted in yellow).

Constituent	R^2	PPCC	B_P	PLR	E
Chloride	88.6	0.985	-8.9	0.911	0.453
SC	94.6	0.981	-3.7	0.963	0.822
TSS	92.5	0.995	237	3.4	-13.1
Sulfate	96.9	0.995	-3.6	0.964	0.848
TP	94.3	0.977	83	1.832	-1.43
TDP	90.4	0.981	60	1.600	-0.611
Particulate-P	94.4	0.997	83	1.829	-1.45
Orthophosphate-P	90.3	0.989	77	1.772	-2.59
SOC-P	91.7	0.989	6.67	1.067	0.790
TN	98.2	0.989	0.3	1.003	0.817
TDN	98.0	0.990	-1.2	0.988	0.937
Particulate-N	86.7	0.967	232	3.317	-18.5
Nitrate-N (IC)	94.4	0.997	-7.5	0.925	0.769
Nitrate-N (FIA)	94.3	0.998	-7.8	0.922	0.770
Nitrite-N	91.6	0.973	7.4	1.074	0.759
Ammonium-N	83.6	0.990	51.3	1.513	-2.17
DON	92.7	0.992	19.9	1.199	0.860